# In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion

Scherer, Klaus R.; Clark-Polner, Elizabeth Paige; Mortillaro, Marcello

**How to cite**

SCHERER, Klaus R., CLARK-POLNER, Elizabeth Paige, MORTILLARO, Marcello. In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. In: International Journal of Psychology, 2011, vol. 46, n° 6, p. 401–435. doi: 10.1080/00207594.2011.626049

This publication URL: https://archive-ouverte.unige.ch/unige:96441

Publication DOI: 10.1080/00207594.2011.626049

INTERNATIONAL JOURNAL OF PSYCHOLOGY, 2011, 46 (6), 401–435

Psychology Press
Taylor & Francis Group

# In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion

**Klaus R. Scherer, Elizabeth Clark-Polner, and Marcello Mortillaro**

Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Do members of different cultures express (or ‘‘encode’’) emotions in the same fashion? How well can members of distinct cultures recognize (or ‘‘decode’’) each other's emotion expressions? The question of cultural universality versus specificity in emotional expression has been a hot topic of debate for more than half a century, but, despite a sizeable amount of empirical research produced to date, no convincing answers have emerged. We suggest that this unsatisfactory state of affairs is due largely to a lack of concern with the precise mechanisms involved in emotion expression and perception, and propose to use a modified Brunswikian lens model as an appropriate framework for research in this area. On this basis we provide a comprehensive review of the existing literature and point to research paradigms that are likely to provide the evidence required to resolve the debate on universality vs. cultural specificity of emotional expression. Applying this fresh perspective, our analysis reveals that, given the paucity of pertinent data, no firm conclusions can be drawn on actual expression (encoding) patterns across cultures (although there appear to be more similarities than differences), but that there is compelling evidence for intercultural continuity in decoding, or recognition, ability. We also note a growing body of research on the notion of ingroup advantage due to expression ‘‘dialects,’’ above and beyond the general encoding or decoding patterns. We furthermore suggest that these empirical patterns could be explained by both universality in the underlying mechanisms and cultural specificity in the input to, and the regulation of, these expression and perception mechanisms. Overall, more evidence is needed, both to further elucidate these mechanisms and to inventory the patterns of cultural effects. We strongly recommend using more solid conceptual and theoretical perspectives, as well as more ecologically valid approaches, in designing future studies in emotion expression and perception research.

***Keywords:*** Emotion expression; Emotion perception; Universality; Cultural specificity; Ingroup advantage; Dialect theory; Multimodal expression.

Est-ce que les membres de différentes cultures expriment (ou « encodent ») les émotions de la même façon? Jusqu'à quel point les membres de cultures distinctes peuvent reconnaître (ou « décoder ») les expressions d'émotion des uns et des autres? La question de l'universalité culturelle de l'expression des émotions par opposition à leur spécificité a été un objet de débat passionné pendant plus d'un demi-siècle, mais en dépit de nombreuses recherches empiriques jusqu'à ce jour, il n'y a toujours pas de réponse convaincante. Nous soutenons que état de chose insatisfaisant est dû en grande part à un manque de préoccupation à l'égard des mécanismes précis impliqués dans l'expression et la perception de l'émotion et nous proposons d'utiliser une modification du modèle de la lentille de Brunswick en tant que cadre approprié pour la recherche sur cette question. Sur cette base, nous présentons un relevé exhaustif de la documentation disponible et nous soulignons des paradigmes de recherche susceptibles de produire les données nécessaires pour résoudre le débat sur l'universalité de l'expression des émotions par opposition à leur spécificité. Partant de cette nouvelle approche, notre analyse révèle qu'étant

Correspondence should be addressed to Klaus R. Scherer, Swiss Centre for Affective Sciences, University of Geneva, rue des Battoirs 7, CH-1205 Geneva, Switzerland. (E-mail: klaus.scherer@unige.ch).

This research has been partially funded by an ERC Advanced Grant in the European Community's 7th Framework Programme under grant agreement no. 230331-PROPEREMO to the first author. The authors gratefully acknowledge contributions by Tanja Bänziger, Nele Dael, Adam Gardner, Sona Patel, and Vera Sacharin.



http://www.psypress.com/ijp http://dx.doi.org/10.1080/00207594.2011.626049

donné la pauvreté de données pertinentes, il n'est pas possible de tirer de conclusions fermes sur les configurations réelles d'expression (encodage) valables pour toutes les cultures (même s'il semble y avoir plus de similitudes que de différences), mais il y a des données convaincantes en ce qui a trait à la continuité à travers les cultures pour ce qui est de l'aptitude à décoder ou reconnaître. Nous notons aussi un ensemble croissant de recherches sur la notion de l'avantage pour l'endogroupe d'utiliser des « dialectes » d'expression au-delà des configurations générales d'encodage et de décodage. Nous suggérons de plus que ces configurations empiriques peuvent trouver une explication à la fois par l'universalité des mécanismes sous-jacents et la spécificité culturelle dans la formation et la régulation de ces mécanismes d'expression et de perception. Dans l'ensemble, il faut plus de données à la fois pour comprendre ces mécanismes et pou faire l'inventaire des configurations des effets culturels. Nous recommandons fortement d'utiliser des perspectives conceptuelles et théoriques plus solides, de même que des approches plus valides d'un point de vue écologique pour concevoir les études futures concernant la recherche sur l'expression et la perception des émotions.

Expresan o codifican las emociones de la misma manera los individuos de culturas diversas? ¿Con qué precisión pueden las personas de distintas culturas identificar o decodificar las expresiones emocionales del otro? El problema de la universalidad cultural versus la especificidad de las expresiones emocionales ha sido un tema de arduo debate por más de medio siglo y, a pesar de contar con una cantidad importante de investigaciones empíricas hasta el momento, no se han obtenido respuestas convincentes. Se sugiere que este estado insatisfactorio sobre el tema se debe en gran medida a la falta de interés respecto de los mecanismos precisos relacionados con la expresión y percepción emocionales, por lo que se propone utilizar un modelo modificado de lente Brunswikiano como marco de investigación en esta área. Sobre esta base se brinda una revisión exhaustiva de las publicaciones existentes, y se apunta a paradigmas de investigación que probablemente proporcionen la evidencia necesaria para resolver el debate sobre la universalidad o especificidad cultural de las expresiones emocionales. Al aplicar esta perspectiva novedosa, el análisis revela que, dada la escasez de datos pertinentes, no se pueden sacar conclusiones sólidas sobre patrones de expresión (codificación) transculturales (aunque aparentemente se observan mayores similitudes que diferencias), pero que existen evidencias convincentes de continuidad intercultural en lo que se refiere a la habilidad de codificación o reconocimiento. Al mismo tiempo, se observa un cuerpo creciente de investigaciones sobre la noción de las ventajas de estar dentro de un grupo determinado, debido a las expresiones del "dialecto", que están más allá de la codificación general o de los patrones de codificación. Asimismo, se sugiere que estos patrones empíricos podrían ser explicados tanto por la universalidad de los mecanismos subyacentes como por la especificidad cultural en los aportes y la regulación de estos mecanismos de expresión y percepción. En general, se requieren mayores evidencias, tanto para aclarar más estos mecanismos como para hacer un inventario de los patrones de los efectos culturales. Se hace una fuerte recomendación de usar perspectivas conceptuales y teóricas más sólidas, como también enfoques ecológicamente más válidos para el diseño futuro de estudios sobre expresión y percepción emocionales.

Scholars have been debating the nature and function of emotion and expression for millennia. Within the past 200 years, the focus has turned to trying to characterize these phenomena *empirically*, using rigorous scientific methods. Recently—within the past few decades—there has emerged a subset of research dedicated to examining how these phenomena may be influenced by culture. Here, we review the literature generated to date on emotion expression and perception with a special emphasis on crosscultural work, and make recommendations as to how best to interpret the current evidence, as well as how to design future studies to advance this popular, but still maturing, area of study.

The question of whether the expression of emotion is universal across peoples or culturally specific is pertinent for debates concerning the definition of emotion itself. If emotions were indeed expressed and interpreted differently across cultures, the nature and determinants of the emotion process would need to be re-examined for each cultural group. For students of emotion it is thus essential to follow the research in this domain and to evaluate it critically.

A cursory examination of the empirical evidence now accumulated reveals no clear-cut answer as to whether emotion is expressed in the same way throughout the world, or differentially by culture. Most of the studies conducted to date have used variations on a single paradigm—in which expressions of a person of one culture are shown to members of either the same or a different culture, whose task is to identify the underlying emotion. These studies reliably report no great variation in the rates with which people can make this identification accurately, though there are some (seemingly small) differences. These differences show only few discernible patterns (save for a within-culture advantage found in several studies), although a serious lack of systematic variation in expresser–perceiver culture (within individual studies, and within the literature as a whole) precludes any conclusions at this point in time.

A more careful consideration of this literature reveals that the findings produced by this body of research may be misleading. As noted, much of the research on culture and emotion expression conducted to date has utilized a single experimental paradigm in which expressions of emotion are characterized by the way in which an *observer* labels them. In other words, these studies have confounded emotion expression and emotion *perception*.

That emotion expression and perception have been lumped together is almost to be expected. After all, it is only by means of perception that we are aware of expression in the first place. Experientially, the two may be indistinguishable. Conceptually and experimentally, however, they can and must be disentangled if we are to understand the underlying processes. The consequence of the current confound is that we cannot, with confidence, conclude from the findings of these existing studies much at all about the influence of culture on the *expressions* produced by an individual when he or she is experiencing emotion, nor can we form firm conclusions about potential cultural influences on the *perception* of those expressions by others.

This review will therefore cover only briefly the conclusions drawn within the existing literature. We will instead focus on the empirical evidence itself. We will propose a conceptual framework by which this existing evidence may be understood, and which may improve our ability to find culturally specific patterns within the data, where they exist, and reorganize the existing data accordingly.

First, however, we provide a brief description of the history of research on emotion expression and the possible effects of cultural differences. We then briefly review the conclusions drawn by the existing studies on emotion and expression, stressing that we believe them to lack a solid foundation (due to the confound in the sole study design used; see Shadish, Cook, & Campbell, 2002 for a discussion of mono-operation and monomethod biases.)

Next, we describe a conceptual model for emotion expression, which may help to guard against confusion of the emotion expression and perception processes. As there are parts of this model for which there are no culture-specific data (the existing studies having been limited to the recognition design already noted), we will also cover in this section findings that do not speak directly to potential cultural differences, but may help elucidate the emotion expression/perception mechanisms, and which may thus help to guide future studies that *do* look specifically at how culture may influence these processes. Finally, having outlined the components of the emotion expression/perception mechanism, we discuss some potential pathways by which cultural influence may operate, in those cases where culturally specific patterns of emotion expression *are* found.

## HISTORICAL BACKGROUND

That emotion expressions were universal was long the assumption. Neither the philosophers who pondered emotion in antiquity nor the pioneers of the more recent scientific study of these phenomena explicitly considered the idea that the mechanism by which emotion is expressed may be influenced by, or incorporate, learned (read: culturally specific) elements.

The first explicit formulation of such a universalist position must be credited to Charles Darwin. Having based his theory of evolution by natural selection on phylogenic continuity in morphological features, Darwin set out to buttress his argument by demonstrating the same continuity in behavioral traits. He chose the expression of emotion as an example behavior, and wrote to his correspondents around the world to solicit descriptions of local expression variants in both man and animal. Based on this material, Darwin posited both interspecies and *intra*species continuity of expression. As an exemplar, he pointed to cross-cultural continuity in human expression, writing: "the different races of man express their emotions and sensations with remarkable uniformity throughout the world" (Darwin, 1872/1998, pp. 130–1).[1]

Although Darwin's focus on expression would not catch on for another few decades, early emotion theorists were influenced by his suggestion that emotion might be rooted in basic biological systems, and their work thus similarly (though less explicitly) supported the view that the expression of emotion was likely to be universal. Thus, trying to develop a universal theory of emotion, William James asserted that emotion *is* the subjective experience of physiological changes within our bodies (James, 1884). If emotion were

[1] Konrad Lorenz and Niko Tinbergen would later win a Nobel Prize (1973) for continuing this line of research, extending the concepts of analogy and homology beyond morphology to behavior.

emergent from very *basic* biological processes it should be uniform within the human species, and so should be emotion expression.[2]

By the beginning of the twentieth century, the idea of biologically based emotions had entrenched itself. It became so easily accepted that the idea that emotion (and emotion expression) should be relatively universal began to gain explicit recognition. Silvan Tomkins, in his Neo-Darwinian theory of emotion, for example, assumed that we share affect states with other animals and that human expressions are shared over cultures (Tomkins, 1962). A group of researchers influenced by Tomkins, in particular Paul Ekman and Carroll Izard, became the first to demonstrate the universality of emotion expression *empirically*, reporting that individuals from different cultures could recognize each other's expressions of emotion with much greater accuracy than would be expected by chance (Ekman, Sorenson, & Friesen, 1969; Izard, 1971).

On the theoretical front, however, the tide had already begun to change. Whereas theorists within psychology had previously proposed that emotion was primarily the experience of rudimentary physiological sensations (James, 1884) or of basic emotion programs (Ekman, 1992a; Izard, 1971; Tomkins, 1962), some were now suggesting the existence of a cognitive appraisal component (Arnold, 1960; Lazarus, 1968; Lazarus, Coyne, & Folkman, 1984). Cognitive ability is more complex, and more recently developed, than the basic physiological and proprioceptive processes that had earlier been thought to form emotion. It should thus also be more phylogenetically exclusive. Therefore, theories that rely more heavily on cognitive components should predict less interspecies continuity in emotion and emotion expression than those that focus on basic psychobiological factors. But whereas there is much variation in cognitive ability between species, there might be relatively little variation within species. Furthermore, it is questionable to what extent intraspecies variation is tied to culture.[3] The notion that cognitive appraisal processes play a major role in the elicitation and differentiation has become increasingly popular (see Ellsworth & Scherer, 2003). This suggests the possibility that there is some degree of emotion specificity with respect to species and cultures, given differences in underlying appraisal patterns, for example due to appraisal biases produced by culture-specific value systems (Scherer & Brosch, 2009).

In general, however, the efforts of biologists and psychologists are generally directed toward finding *similarities* in the mechanisms underlying human behavior, while those of anthropologists and ethnologists tend to be directed toward the discovery of cultural *differences*. Indeed, anthropological and ethnographic explorations of potential crosscultural disparities in thought, language, and customs produced the first detailed paradigms of potential culture bounding of emotion expression (see Lutz & White, 1986, for an overview). Among the most well-known of these paradigms are those posited by Ray Birdwhistell and Margaret Mead. Birdwhistell argued that emotion expression, like language, should be understood as a collection of sign structures, which are learned and thus culturally bound (Barfield, 1997; see also the account given by Ekman in the afterword to his new edition of Darwin, 1872/1998). Mead—a critic of the biologists' and psychologists' commonality-based approach—suggested that, in different cultures, different expressions were more commonly utilized during different developmental periods (Mead, 1928).

Despite the increasingly nuanced theoretical perspective, however, empirical work continued to be carried out almost exclusively in the universalist tradition. In fact, a preponderance of the empirical work completed to date has the identical paradigm as studies that were seminal to this area of research 40 years ago. With this empirical predominance came what appeared to be widespread acceptance of a universalist view, dominating textbook coverage of emotion until very recently. Research on potential cultural influence on emotion expression continued to be somewhat biased toward finding similarities, not differences.

Within the past decade, however, perspectives have begun to change. Researchers have recently started to explore the idea that emotion, and specifically the expression of emotion, may be more complex than commonly realized. Today, the

[2] This is assuming, of course, that expressed emotion reliably reflects experienced emotion. Fridlund (1994), putting forth a behavioral ecology view, has posited that there may be little connection, but no empirical evidence supporting this claim has been published so far.

[3] There is a distinction to be made here between cognitive ability and patterns of specific cognitions. Whereas there is no evidence that cultures differ in ability, there is evidence that people from different cultures may think in different ways (Mesquita, Frijda, & Scherer, 1997). We make the same argument in this review for the expression of emotion: there is universality in ability, but people from different cultures learn to deploy that ability in slightly different ways.

literature on emotion expression and culture contains what could be construed as support for both sides of the debate (see contributions in Ekman & Rosenberg, 2005; Russell & Fernandez-Dols, 1997). Little effort has been made, though, to reconcile these findings—for and against universality—and to elucidate the middle ground. In addition, a more sophisticated theoretical framework will be necessary to guide future research. In this spirit, we will first examine the link between expression and perception (and inference).

## EXPRESSION AND IMPRESSION: THE BRUNSWIKIAN LENS MODEL

### A common confound

When we speak of *the expression of emotion*, we usually mean something more than what the words, strictly speaking, imply. The word *expression* would seem to refer only to the process by which a particular set of features is produced (the word *expression* itself is derived from the Latin *ex*, "out," and *pressio*, "pressing, pushing"), and/or to the features themselves "pressed out." However, much of the work referred to under the label of emotion expression in reality studies the perception of emotion expression. Emotion expression is thus commonly conflated with emotion *perception*.

How did this happen? Emotion expression research has traditionally taken one of two forms, the "production" study and the "recognition" study. In production studies, expressions are analyzed by measuring objective parameters such as facial muscle movements or acoustic parameters of the voice. In recognition studies, expressions are judged by observers, who are shown expressions and asked to identify the underlying emotion(s). This type of study is epitomized by the early work of Ekman et al. (1969) and Izard (1971).

Both types of study purport to measure expression. In juxtaposition, however, it is obvious that they tap distinct constructs; production studies analyze the characteristics of that which is (ex)pressed out, whereas recognition studies examine the ability of an observer to identify emotions on the basis of their perception of the expressed features and the inferences they draw from them. Clearly, both production and perception/inference are part and parcel of the process of communicating emotions—we do not commonly think of an emotion as being communicated until someone else becomes aware of it—but they are *different* parts. This distinction, however, is lost when we speak only of expression (including recognition) without specifying the respective mechanism. By overlooking the distinction, we may be hampering our ability to find patterns in the data that allow us to estimate the degree to which the expression or the recognition of emotion may be universal or culturally specific, respectively. To be more precise, there might be cultural differences in the way expressions for different emotions are produced (resulting in different facial or vocal expression feature configurations). Or there could be differences in the way members of different cultures perceive these features and the inferences they draw from them in attempting to recognize the underlying emotion. Or there could be cultural differences in both processes.

### A new conceptual approach: The Brunswikian lens model

To remedy this unsatisfactory state, we suggest that the mechanisms involved, and the available data, may be best understood from the perspective of a modified version of Brunswik's functional lens model of perception (Brunswik, 1952, 1956; Gifford, 1994; Hammond & Stewart, 2001). The detailed argument for the modification of the model can be found elsewhere (Scherer, 1978, 2003; see also Kappas, Hess, & Scherer, 1991; Scherer, Johnstone, & Klasmeyer, 2003), and so we will only briefly outline the latest version here.

The modified Brunswikian lens model is shown in Figure 1. It proposes that the communication of information concerning the emotional state of an individual comprises two distinct, but interrelated, processes—expression and perception/inference. We propose that the clear conceptual modeling of this distinction will benefit future research.

According to the model, the expression of the internal emotion process consists of an externalization in the form of physical signals in the face, the voice, and the body, such that it becomes perceptible to another. We call this the "encoding" process. The specific way in which emotions are reflected physically and the mechanisms involved depend on the specific modality, but the underlying concept is the same in each. The basis of any functionally valid communication of emotion via motor expression is that different types of emotion are characterized by *unique patterns of bodily, facial, or vocal cues*. Without distinct patterning, the nature of the underlying encoder state could not be communicated reliably. We call this patterning of motor behavior "expressed cues."

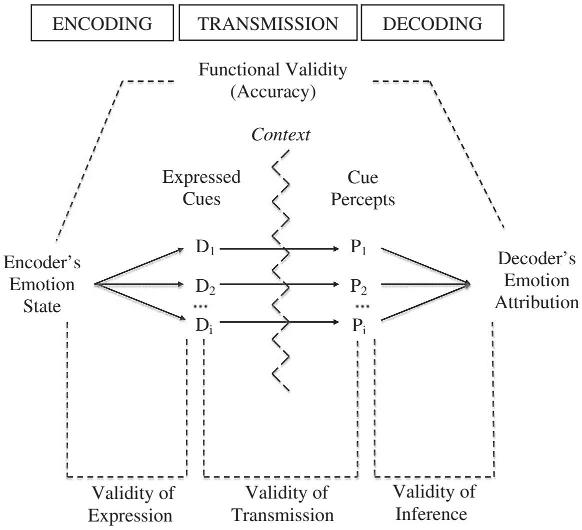


**Figure 1.** Modified Brunswikian lens model, applied to the process of the expression and perception of emotion. See detailed explanation in the text (adapted from Scherer, Johnstone, & Klasmeyer, 2003, p. 434 by permission of Oxford University Press).

In vocal expression, for example, the emotion of the speaker may be accompanied by physiological changes, which will affect the vocal musculature involved in respiration, phonation, and articulation. In facial expression, the patterning consists of an activation of facial muscle configurations (e.g., around mouth, nose, and eyes) that change facial appearance in manifold ways, including the production of wrinkles. If there is sufficiently specific patterning of these vocal–acoustic parameters or facial features, and if specific patterns are reliably related to different aspects of the unfolding emotion process, an observer can use these cues to infer the emotional quality of the underlying state (see Scherer, 1986; Scherer & Ellgring, 2007a, for detailed descriptions).

The patterning of expressive features is measured by the type of research we termed "production studies." These studies inventory the characteristics of potential channels of expression (voice, face, body, etc.), and how they may change during the experience of different emotions. If the patterning of these changes were reliably similar within cultures, but distinct between them, a cultural specificity in this aspect of expression would exist.

Specificity in one aspect—the encoding, for example—does not necessarily imply specificity overall, however: Potentially communicative cues—expressed cues or "distal cues"[4]—may be produced by the emoter, but they must be "correctly" perceived and interpreted in order to provide information to the observer. In the model, the cues as perceived by the observer are called "cue percepts" or "proximal cues."[5] The process by which distal expressive behaviors become available for perception is called "transmission." Exactly how transmission occurs depends on the modality of expression (e.g., visual, auditory). The transmission of vocal cues, for example, occurs via sound waves captured by the auditory perceptual mechanisms of the observer (see Scherer, 2003).

On the basis of these cue percepts, the observer or perceiver *infers* what emotion(s) the emoter is experiencing. Brunswik called this process *cue utilization*. Where the proximal cues are indeterminate, contextual information may be utilized to disambiguate the inferred emotion state. In our model, we term this inference process (from cues and/or context) "decoding." The studies that we called "recognition studies" above do not allow to understand how the expressed cues are perceived and what inferences are drawn from them. We need both production studies (to measure the patterns of expressed cues) and perception/inference studies (to assess the observers' emotion attributions from cue percepts). And only when production and perception/inference are jointly measured in the same study, for the same set of expressions, can these two processes be untangled.

## The utility of the model: Functional validity and experimental accuracy

Brunswik used the term "functional validity" to refer to what he called the "achievement" or "efficiency" of the model, in our case the efficiency with which valid information about an individual's emotional state can be communicated to an observer—or to put it more simply, how well the observer can recognize the expressed emotion. This is of course what most studies reviewed below attempt to study. However, the lens model allows the different parts of the overall mechanism to be formalized and experimentally measured. In the model shown in Figure 1, we distinguish between three types of validity: (1) the validity of

[4] In the Brunswikian tradition, "distal" is used in the sense of remote or distant from the observer. We also refer to distal cues as "expressed cues" to highlight the fact that they are produced by the expresser.

[5] "Proximal" is meant in the sense of closeness to the observer. We also use "cue percepts" in recognition of the fact that we are examining the process from the perspective of both perceiver and emoter.

expression—this refers to the specificity of the expressed cues and cue configurations ($D_1 \ldots D_i$) for specific emotions. If expressed cues do not vary between emotions, no information on emotional quality is transmitted; (2) the validity of transmission—this refers to the reliable transfer of the cue characteristics or signals to the sensory organs of the perceiver, thus providing a reasonably precise mapping of the distal cues D into the cue percepts or proximal cues ($P_1 \ldots P_i$). Thus, if noise drowns out central acoustic parameters of a vocal expression, or if the perceiver is deaf, there is no valid transmission; and (3) the validity of inference—this refers to the match between expression patterning and inference rules. If a certain configuration of specific cues is a valid indicator of emotion X and the observer, having perfectly well perceived the configuration, infers emotion Y, there can be no valid attribution. Validity is not an all-or-nothing phenomenon but continuously varies from low to high. Obviously, given the interdependence of the validities, the overall functional validity of the whole process can only be high if all partial validities are also reasonably high.

Overall functional validity can be operationalized through empirically measured accuracy, which can be done, following Brunswik's recommendation, by correlating the real state with the attributed state. In the case of emotion recognition this is difficult, as generally only categorical judgments are made. In this case accuracy is measured as the percentage of accurate attributions (the decoded emotion) given a categorical criterion (the encoded emotion). While this assessment is generally done without recourse to the model, the latter, based on the measurements of expressed cues and cue percepts, can explain why functional validity is high or low. For example, where fit is lacking, one can immediately identify the possibility that the respective emotional state does not produce reliable externalizations in the form of specific distal cues in the face or voice. Alternatively, valid distal cues might be degraded or modified during transmission and perception in such a fashion that they no longer carry the essential information when they are proximally represented in the observer. Similarly, it is possible that the proximal cues reliably map the valid distal cues but that the inference mechanism—that is, the cognitive representation of the underlying relationships—is flawed in the respective listener (e.g., from lack of sufficient exposure or inaccurate stereotypes). Clearly, this approach is of great utility in the case of the question concerning universality or cultural specificity of emotion communication—it would allow determining with great precision if, where, and to what extent there are culture specificities in the process.

## A Brunswikian perspective on the existing literature

Unfortunately, we have not been able to identify many published studies that have adopted the lens model approach in the standard emotion expression literature. Thus it remains for future research to provide a detailed assessment of the parts of the expression–perception process in which culture plays a major role. However, we have decided to use the model to organize our review of the literature. A conceptual approach such as this demands a different perspective on the data; we must distinguish between data that speak to encoding and data that speak to decoding, and apply our inferences from these distinct data sets strictly to the appropriate *sub*components of the overall emotion expression–perception process. In general, a review of the literature reveals:

(1) a (relatively) small number of studies that measure the encoding of emotion (*production studies*), or studies in which actor-portrayed, induced, or observed emotion expressions are analyzed with respect to objectively measured distal indicators or cues (e.g., facial action units [AUs] or acoustic parameters).
(2) very few *cue manipulation studies*, in which facial or vocal cues are experimentally manipulated (e.g., through facial or vocal synthesis), which allow us to examine the way in which expressed cues are used in observer inference (but see Breitenstein, Van Lancker, & Daum, 2001; Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Murray & Arnott, 1993; Scherer, 2003; Scherer & Oshinsky, 1977; Wehrle, Kaiser, Schmidt, & Scherer, 2000).
(3) almost no *transmission and representation studies*, which look at the process by which expressed cues become cue percepts
(4) a large number of decoding studies—in the very few cases where these studies measured both perceived emotion and objective characteristics of the expression perceived, they may speak to both the encoding and decoding phases of the model (true expression and perception); where they measured only observer perception, no such inference can be made for either process

(5) only a few empirical studies that have sought to generate data for one or even several components of the model (see Gifford, 1994; Juslin, 2000; Scherer, 1978; for the rare attempts).

It is thus to be expected that many of the existing data (i.e., decoding studies) will do little to expand our understanding of the overall emotion expression–perception process until we have complementary empirical analyses of the other phases. Future research should be designed with this in mind.

## REVIEW OF THE EXISTING EMPIRICAL EVIDENCE

The current literature comprises half a century's empirical investigations of the emotion expression-perception process. We will first define characteristics by which these investigations may be described, and then review their findings, guided by the Brunswikian lens model.

### Dimensions along which studies vary

#### *Modality*

Emotion can be expressed through multiple channels. The vast majority of all expression research studies, however, have been done on facial and vocal expression, and this holds for crosscultural research as well. Here we address crosscultural research on expression in the face and in the voice and briefly note the small number of additional studies on gesture and posture.

#### *Expression elicitor*

In order to study the crosscultural nature of emotion expression, one needs expressions of emotion as experimental stimuli. Emotion—and thus expression—can be elicited in multiple ways. Expression, however, can also be feigned, or enacted. Many different methods can be used to elicit portrayals. Those used in the extant literature can be classified into three major categories: natural, induced, and simulated emotional expression.

*Natural expression.* Work in this area has made use of material that was recorded during naturally occurring emotional states of various sorts, such as dangerous flight situations for pilots, journalists reporting emotion-eliciting events, affectively loaded therapy sessions, or talk and game shows on TV (see the reviews on corpora by Cowie et al., 2010; on facial expression studies by Ekman & Rosenberg, 2005; and on vocal expression studies by Scherer, 2003). For example, the use of naturally occurring voice changes in emotionally charged situations, recorded on the fly, seems the ideal research paradigm because it has very high ecological validity. However, there are some serious methodological problems. Voice samples obtained in natural situations, often only for a single speaker or a very small number of speakers, are generally very brief and frequently suffer from bad recording quality. In addition, determining the precise nature of the underlying emotion and the effect of regulation is problematic (see following paragraphs).

*Induced emotions.* Another way to study emotional expression is to experimentally induce specific emotional states in individuals and then record their expressive behavior. Most induction studies have used indirect paradigms that include stress induction via difficult tasks to be completed under time pressure, the presentation of emotion-inducing films or slides, or imagery methods (see the reviews on facial expression studies by Ekman & Rosenberg, 2005, and on vocal expression studies by Scherer, 2003). Although this approach—generally favored by experimental psychologists because of the degree of control it affords—does result in comparable expression samples for all participants, there are a number of serious drawbacks. However, these procedures often produce only relatively weak affect (e.g., Laukka, Neiberg, Forsell, Karlsson, & Elenius, 2011; Tcherkassof, Bollon, Dubois, Pansu, & Adam, 2007). In addition, despite using the same procedure for all participants, one cannot necessarily assume that exactly the same emotional states are produced in all individuals, precisely because of the individual differences in event appraisal mentioned earlier. Scherer and his collaborators, in the context of a large-scale study on emotion effects in automatic speaker verification, have attempted to remedy some of these shortcomings by developing a computerized induction battery for a variety of different states (Scherer, Johnstone, Klasmeyer, & Bänziger, 2000). Similarly, Tcherkassof and colleagues (2007) used computer tasks to induce mild positive (interest and amusement) and negative states (irritation and worry).

*Simulated (posed, portrayed, enacted) expression.* This family of methods has been the preferred way of obtaining emotional expression samples in this field. Professional or lay actors are asked to produce facial and/or vocal expressions of emotion (often using standard verbal content) based on emotion labels and/or typical scenarios (see, e.g., Banse & Scherer, 1996; Bänziger, Mortillaro, & Scherer, in press; Hawk, Van Kleef, Fischer, & Van der Schalk, 2009; Wallbott & Scherer, 1986). It should be noted that there are important differences in ecological validity between *posing* (producing an iconic expression based on an emotion label and following precise instructions on which movements to produce), *portraying* (producing an expression that is typical for a given scenario), or *enacting* (producing an emotion expression based on reliving an appropriate emotion experience of one's own; Scherer & Bänziger, 2010). There can be little doubt that this approach yields much more intense, prototypical expressions than are found in induced states or even natural emotions (as the latter are likely to be highly controlled; see Scherer & Bänziger, 2010). However, actors may overemphasize relatively obvious cues and miss subtler ones that might appear during the natural expression of emotion. Unfortunately, there are few empirical studies comparing portrayals to induced emotion expressions that could provide evidence for this notion. In a recent study on vocal expression, acted happy and sad expressions were compared to happy and sad moods induced by the Velten induction procedure for the same speakers. The data showed that the two elicitation procedures produced very similar differences in vocal parameters between happy and sad states, suggesting that it may not matter very much whether emotional expressions are enacted or experimentally induced, at least for some major emotions (Scherer, 2011).

It has often been argued that emotion portrayals reflect sociocultural norms or expectations more than the psychophysiological effects that occur under natural conditions. However, it can be argued that all publicly observable expressions are to some extent "portrayals" (given the social constraints on expression and unconscious tendencies toward self-presentation; see Scherer & Bänziger, 2010). Furthermore, expression portrayals are reliably recognized by observer judges (see below), and so it can be assumed that they reflect at least in part "normal" expression patterns (if the two were to diverge too much, the acted version would lose its credibility). It cannot be denied that actors' portrayals are influenced to some degree by conventionalized stereotypes of expression, however. To obtain a maximal level of authenticity in the expressions of actors, one may use the enacting procedure whereby the actors are asked to use the techniques pioneered by Stanislavski (Roach, 1985) to produce an appropriate internal feeling through recourse to personal memory or mental imagery (Scherer & Bänziger, 2010).

*Overall*. Each of the methods that have been used to obtain emotion expression samples has both advantages and disadvantages. In the long run, the best strategy is likely to look for convergences between all three approaches in the results. We will now summarize the evidence, mostly obtained in portrayal studies, according to expression modality, noting how the expressions studied were elicited so as to facilitate comparison and encourage further research. Unfortunately, the number of studies allowing a genuine comparison across cultures is extremely small. In the review below, we present the studies that have been done, and make a comparison between Western and non-Western cultures (Western comprising Europe, North America, and South America, as well as Australia.). Whereas potential cultural differences may not cluster in this (Western/non-Western) manner, these are the categories that have traditionally been used to classify cultures within this research. This binary distinction is unsatisfactory, as the criteria for defining Western and non-Western cultures are clearly very difficult to specify and as it is often not clear how "Westernized" some of the participants in big cities of non-Western cultures are. Furthermore, there is clearly quite a lot of variance within each group. Unfortunately, this variance is difficult to estimate as there is often no replication for a particular country. Ideally, one would hope that at some point there will be a sufficiently large data set to correlate the results of studies on emotion expression and recognition with established indicators of country differences with respect to important variables (such as predominant values, but also climatic, geopolitical, economic, and religious indicators; see Scherer, 1997).

We organize our review by presenting the major work for each of the central expression modalities (facial, vocal, gestural/postural) separately for encoding and decoding studies. For the decoding work we present and comment on the average recognition accuracy for a set of basic emotions that are shared across studies. For the convenience of the reader we provide a list of all studies considered in this review, along with pertinent information on encoder and decoder culture as well as the number of emotions studied, in the Appendix.

## EMPIRICAL FINDINGS

### Facial expression

#### *Facial encoding studies*

The earliest record of a pictorial representation of facial expressions dates back almost eight millennia (as shown in the enigmatic expressions of the Ain Ghazal statues found near Amman, Jordan, which date to about 7500 BC; Kleiner & Mamiya, 2006). Since then, painters and sculptors have produced a massive amount of work showing highly differentiated facial expressions. The invention of photography further facilitated the documentation of facial expressions, and with the recent development of digital photography and video recording (including portable telephone capture), the amount of material potentially available for studying expressive encoding is staggering. However, it is difficult to objectively define what emotion is present in any one depiction or portrayal. Evaluation by an observer confounds expression and perception, and expressions produced with a specific emotion in mind are also problematic, as already discussed.

The Facial Action Coding System (FACS), developed by Ekman and Friesen (1978; Ekman, Friesen, & Tomkins, 1971) and based on prior work by Hjortsjö (1970), has become the method of choice to analyze the distal cues associated with the facial expressions of emotions; FACS allows description of facial expressions in terms of the configuration of individual facial AUs (minimal units of discrete changes in facial expression that are due to the innervations of one or several muscles). These units are free of inference and coded only on the basis of their individual occurrence; thus, they can be used to code any changes in facial expression, independently of the underlying causes. For emotional expression, Ekman and colleagues (Ekman & Friesen, 1975; Ekman, Friesen, & Hager, 2002) have, over the course of many years, developed a comprehensive set of predictions about the types of AU configuration to be expected for a set of basic emotions (based on theoretical considerations and systematic observation.)

The use of FACS to code AUs is extremely time-consuming, and thus the technique has been used mainly to study specific facial behaviors and emotions (e.g., Reeve & Nix, 1997) and in a limited number of sophisticated research studies. A contributed volume edited by Ekman and Rosenberg (2005) brings together many of the pertinent studies. Unfortunately, at the time of this writing, no systematic effort has been made to inventory the results of these studies with respect to the empirically found AU configurations for the typical set of basic emotions and to compare the results to predicted patterns.

There are, however, some general observations to be made. Many studies do not find the complete set of AU configurations predicted by Ekman and colleagues (Ekman & Friesen, 1975; Ekman et al., 2002) for specific emotions. This observation holds even in the case of infants and young children (Camras et al. 2002; Camras, Bakeman, Chen, Norris, & Cain, 2006; Scherer, Zentner, & Stern, 2004), where one would expect little regulation or strategic display. Ekman (2003) has acknowledged this frequent absence of complete emotion-specific configurations. He reconciles this finding with theoretical predictions by allowing for affect programs to be only partially active at any one time, which he posits would be reflected in the activation of a subset of the predicted AU patterning. Similarly, Matsumoto and colleagues commented that "spontaneous expression includes not only facial cues that signal emotion, but also extraneous muscle movements that reduce emotion signal clarity" (Matsumoto, Olide, Schug, Willingham, & Callan, 2009, p. 235).

Scherer and Ellgring (2007a) reported results from an enacting portrayal study in which they examined the frequency with which AUs were activated in professional actors' portrayals of major emotions. They found very few of the predicted configurations. It is notable that despite the absence of the complete or even partial patterns of prototypical AU configurations postulated by Ekman and his collaborators for basic emotions, the respective portrayals were generally recognized with rather high accuracy by raters. We suggest that this is evidence for the existence of a sufficient number of valid expressed cues for certain emotions even if the AU configurations specified by affect program theories are not fully realized. We will illustrate this with an example. Table 1 displays the central empirical results of the study performed by Scherer and Ellgring (2007a) for selected AUs, along with the predictions made by an appraisal theory (the componential process model; Scherer, 2001). This theory postulates that facial movements are produced by specific appraisal results and consequent action tendencies, in a sequential, cumulative manner. The authors suggest that the pattern of findings can be interpreted, using plausibility criteria, as supporting this notion and may be more parsimonious than the assumption that there are emotion-specific neuromotor

**TABLE 1**

Theoretically postulated and empirically found patterning of action unit activation in actors' portrayals of emotion expressions (adapted from Scherer & Ellgring, 2007a, p. 121)

| *Action units* | *Hot Anger* | *Cold Anger* | *Panic Fear* | *Anxiety* | *Despair* | *Sadness* | *Elated Joy* | *Happiness* | *Disgust* |
|---|---|---|---|---|---|---|---|---|---|
| Inner brow raiser | 0.31 | 0.31 | 0.94 | 0.63 | 0.88 | 0.56 | 0.44 | 0.31 | 0.25 |
| Outer brow raiser | 0.44 | 0.38 | 0.69 | 0.38 | 0.44 | 0.31 | 0.44 | 0.38 | – |
| Brow lowerer | 0.25 | 0.31 | 0.69 | 0.56 | 0.94 | 0.63 | – | – | 0.56 |
| Upper lid raiser | 0.31 | – | 0.50 | 0.19 | 0.25 | – | 0.19 | – | – |
| Cheek raiser | 0.13 | – | – | – | 0.44 | – | 0.63 | 0.81 | 0.38 |
| Lid tightener | – | 0.19 | – | – | – | – | – | – | – |
| Nose wrinkler | 0.13 | – | – | – | – | – | – | – | 0.13 |
| Upper lip raiser | 0.13 | 0.19 | 0.19 | – | 0.19 | – | – | – | 0.81 |
| Lip corner puller | – | – | – | – | – | – | 0.94 | 0.88 | – |
| Lip corner depressor | – | – | – | – | – | 0.13 | – | – | 0.19 |
| Chin raiser | – | – | – | – | 0.19 | – | – | 0.13 | 0.19 |
| Lip stretcher | 0.44 | 0.25 | 0.25 | 0.19 | 0.19 | – | – | – | 0.25 |
| Lip tightener | 0.19 | 0.19 | – | – | – | – | – | – | – |
| Lips part | 0.13 | 0.31 | 0.13 | 0.19 | 0.25 | 0.50 | – | 0.31 | 0.44 |
| Jaw drop | 0.38 | 0.56 | 0.56 | 0.63 | 0.44 | 0.19 | 0.81 | 0.56 | 0.38 |

Numerical data indicate the proportion of cases in which actors displayed activation for each action unit or emotion combination. A dash indicates that the respective AU was not used by any actor. Shading indicates theoretically predicted patterns of activation: light shading = increased activation; dark shading = greatly increased activation (as compared with baseline).

affect programs as postulated by discrete or basic emotion theories.

### *Facial decoding studies*

It is in the domain of facial expression that the most extensive work on decoding ability has been carried out. Following Tomkins (1962), both Ekman (1972, 1992b) and Izard (1971) and their collaborators, as well as other investigators, conducted studies using standardized series of photographs of theoretically defined expressions of (discrete) emotions (Beaupré & Hess, 2005; Ekman, 1989; Ekman & Friesen, 1971; Ekman et al., 1969; Izard, 1994; for a description of discrete emotions, see Ekman, 1992a). In these studies, photographs are shown to respondents, who are asked to indicate which emotion, from a pre-established list, is being portrayed. The results show that these expressions are, in fact, recognized both within and across cultures with a degree of accuracy significantly greater than would be expected by chance. The vast majority of these studies utilized static photographs as their stimuli. More recently, however, a number of studies have used dynamic portrayals (video recordings) to examine recognition accuracy (Ambadar, Schooler, & Cohn, 2005; Bänziger, Grandjean, & Scherer, 2009; Bänziger et al., in press; Hawk et al., 2009; Simon, Craig, Gosselin, Belin, & Rainville, 2008). As with the static stimuli studies, the accuracy with which participants are able to identify dynamic portrayals is much higher than would be expected by chance. A summary of the findings, separately for static and dynamic samples and for Western and non-Western cultures, is given in Table 2 (facial expression).

Table 2 is organized in such a way as to allow comparison between recognition accuracy percentages, depending on the nature of the encoder and decoder groups, as this will enable evaluation of the notion of ingroup advantage (greater accuracy for decoding expressions of members of one's own social group; see below for a more detailed discussion) for different emotions (to evaluate the relative difficulty of recognizing specific emotions in the face) and for different types of stimulus material (static photo, dynamic video, vocal expressions, and multimodal videos). Table 2 also demonstrates that there are areas for which we found no data in our literature search. Consequently, only some issues can be discussed at the current time.

With regard to facial expressions, the ingroup advantage hypothesis can be examined only for static photos. Here, there is a clear ingroup advantage for Western decoders when they are rating expressions portrayed by Western as compared with non-Western encoders (an overall advantage of about 24%, mostly due to higher recognition accuracy for fear, disgust, and anger). In contrast, surprisingly, for non-Western decoders, there seems to be a Western encoder advantage of about 15% (mostly due to fear,

**TABLE 2**

Percentages of recognition accuracy in crosscultural studies organized by culture (Western versus non-Western) of facial expression encoding and decoding, vocal expression encoding and decoding, and multimodal expression encoding and decoding A)

| | | *Western decoder* | | | | | | | *Non-Western decoder* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* |
| *Facial expression*[a] | | | | | | | | | | | | | | | |
| Western encoder | Face (static) | 91.5 | 82.5 | 73.6 | 73.3 | 73.7 | 72.4 | 77.8 | 86.2 | 77.6 | 68.0 | 56.7 | 64.0 | 61.9 | 69.0 |
| | Face (dynamic) | 65.9 | 50.3 | 58.5 | 70.0 | 67.3 | 63.3 | 62.5. | – | – | – | – | – | – | – |
| Non-Western encoder | Face (static) | 73.4 | 64.3 | 57.9 | 41.3 | 43.9 | 45.0 | 54.3 | 81.0 | 61.1 | 52.7 | 40.0 | 36.1 | 49.7 | 53.4 |
| | Face (dynamic) | 73.0 | 27.0 | 68.0 | 18.0 | 46.0 | 51.0 | 47.2 | – | – | – | – | – | – | – |
| *Vocal expression*[b] | | | | | | | | | | | | | | | |
| Western encoder | Voice | 54.0 | 57.8 | 69.3 | 62.4 | 35.4 | 74.9 | 59.0 | 24.0 | 41.0 | 60.3 | 31.0 | 26.0 | 50.3 | 38.8 |
| Non-Western encoder | Voice | 51.7 | – | 79.3 | 51.7 | – | 66.0 | 62.2 | – | – | – | – | – | – | – |
| *Multimodal expression*[c] | | | | | | | | | | | | | | | |
| Western encoder | Multimodal | 73.0 | 51.7 | 67.3 | 79.3 | 80.0 | 83.3 | 72.4 | – | – | – | – | – | – | – |
| Non-Western encoder | Multimodal | – | – | – | – | – | – | – | – | – | – | – | – | – | – |

A dash indicates that no data are available. [a]Face (static)" = studies using photographs of faces as stimuli; "Face (dynamic)" = studies using videos of facial expressions. The table summarizes data from studies using realistic (not drawn or animated) portrayals of facial expressions of emotion to compare identification accuracy for more than one emotion, using at least two cultural groups (in the case of studies that used photographs), as found in a review of the literature. A complete list of these studies can be found in the Appendix (Tables A1 and A2.) [b]The table summarizes data from studies comparing identification accuracy for vocal portrayals of multiple emotions, as found in a review of the literature. A complete list of these studies can be found in the Appendix (Table A3). [c]No study of multimodal expression utilizing non-Western encoders or decoders was found in a review of the literature. A complete list of the studies summarized in the table can be found in the Appendix (Table A4).

sadness, surprise, and disgust). This may not represent an outgroup advantage; rather, it probably reflects the fact that the stimulus photos are of superior quality because of the long Western research tradition in this area and the existence of well-validated expression corpora. Overall, however, there are no major differences; in particular, the relative success in recognizing specific emotions seems to be quite comparable across the origin of encoders and decoders, suggesting that these differences are due to the functional validity of the expressed cues, the cue percepts, and the inferences for certain emotions. For example, the zygomaticus action (smile) is likely to be a very valid cue for happiness expression and the inference rules for percepts are likely to match this relationship closely.

The difference between static and dynamic displays can be examined only for Western encoder and decoder combinations. Interestingly, generalized across emotion, recognition accuracy appears to be approximately 15% *lower* in the dynamic portrayals (contradicting earlier results with synthetic stimuli; Wehrle et al., 2000). Again, it would be premature to generalize this finding, given that many of the static photo corpora have been intentionally designed to produce highly prototypical expression stills (which can almost be seen to represent iconic representations of specific emotions as used in the form of emoticons), whereas many of the dynamic video corpora have been produced with actors enacting earlier emotion experiences (through Stanislavski's method acting or verbal imagery techniques). Obviously the latter stimuli—in addition to the changing expressions over the course of the video, as well as the much subtler forms of expression, and the possibility of emotion blends—are more difficult to recognize unequivocally. It is worth mentioning that studies using static pictures that were not posed also tend to report lower recognition accuracy than studies using posed facial expressions (e.g., Matsumoto et al., 2009).

Interestingly, whereas the still photo studies show rather higher recognition rates for happiness and surprise than for the other emotions (a standard finding in the area), these differences do not occur with the dynamic video presentations. This could be due to strong effects of zygomaticus action (mouth corners upward) for happiness and wide open eyes for surprise, both cues with strong iconic function, which are very prominent in still photos whereas they may be less domineering in dynamic displays (for example, it is known that the dynamics of a smile can change the meaning attributed to the smile; Ambadar, Cohn & Reed, 2009; Krumhuber & Manstead, 2009).

Appraisal theorists suggested that observers can infer different information from emotional facial expression and this may in turn influence which emotion is inferred. Scherer and Grandjean (2008) hypothesized that observers can use facial actions to recognize appraisal results and the consequent action tendencies, and can thus infer the nature of the expressed emotion. Other scholars compared inferences of emotion labels with social situation antecedents (Haidt & Keltner, 1999) and with emotional action readiness (Tcherkassof & Suremain, 2005) in different cultures. Results showed crosscultural similarities for all types of inference, suggesting the possibility that emotional facial expressions may be able to communicate, reliably and crossculturally, emotion-relevant information that goes beyond a simple emotion label. Future studies should test the types of inference that perceivers can make starting from simple and complex facial action units. As it is now possible to generate experimentally controlled visual stimuli of dynamic facial expressions (e.g., the recently developed FACSGen tool; Krumhuber, Tamarit, Roesch, & Scherer, in press; Roesch et al., 2010)—these tools should be used to systematically investigate the effect of different facial cues on the emotion inference process.

### Voice

*Vocal encoding studies*. Whereas facial expressions have been visually represented over many centuries, sound recordings of vocal emotion expressions have existed only since the invention of Edison's phonograph. Consequently, we have a much smaller corpus with which to work than for facial expression. Despite rapid advances in digital recording and storage (increasingly also possible with portable phones), the situation has not fundamentally changed. In addition to the limitations noted for facial expression, which apply to vocal depictions as well, the need for professional equipment and recording conditions hampers the collection of appropriate stimuli. The set of studies that does exist, however, has yielded some significant insights.

The description of the objective characteristics of vocal expressions (the expressed or distal cues) is usually performed by identifying the acoustic characteristics of the recorded voice signal. Scherer et al. (2003) have reviewed the converging evidence with respect to the acoustic patterns that characterize the vocal expression of major modal

**TABLE 3**

Synthetic overview of selected empirical findings on the effect of emotion on selected vocal parameters (adapted from Scherer et al., 2003, p. 436 by permission of Oxford University Press)

| *Acoustic parameters* | *Happiness/Elation* | *Anger/Rage* | *Sadness* | *Fear/Panic* |
|---|---|---|---|---|
| *Speech rate and fluency* | | | | |
| Number of syllables per second | >= | < > | < | > |
| *Voice source—$F_0$ and prosody* | | | | |
| $F_0$ mean | > | > | < | > |
| $F_0$ deviation | > | > | < | > |
| $F_0$ range | > | > | < | < > |
| $F_0$ final fall: range and gradient[c] | > | > | < | < > |
| *Voice source—vocal effort and type of phonation* | | | | |
| Intensity (dB): mean | >= | > | <= | – |
| Intensity (dB): deviation | > | > | < | – |
| Gradient of intensity rising and falling | >= | > | < | – |
| Relative spectral energy in higher bands | > | > | < | < > |
| Spectral slop | < | < | > | < > |
| Harmonics/noise ratio | > | > | < | < |
| *Voice source—glottal waveform* | | | | |
| Excitation strength (EE) | > | >= | <= | > |
| *Articulation—speed and precision* | | | | |
| Formants—precision of location | = | > | < | <= |
| Formant bandwidth | – | < | > | – |

A dash indicates that no data are available. In specific phonemes, as compared with neutral, $< \approx$ smaller, lower, slower, less, flatter, or narrower; $= \approx$ neutral; $> \approx$ bigger, higher, faster, more, steeper, or broader; $<= =$ both smaller and equal reported; $>= =$ both larger and equal reported; $< > =$ both smaller and bigger have been reported.

emotions. Table 3 summarizes their conclusions in synthetic (qualitative) form. Juslin and Laukka (2003) have provided an additional review, detailing 114 studies, and have reported a meta-analysis of the results. The results of these and future studies that aim to objectively measure distal cues will be of central importance to future research; without having measured concrete acoustic parameters to which a decoder's perception can be compared, we will never be able to disentangle the multiple components of the emotion expression–perception process.

Much of the consistency in the findings is linked to differential levels of arousal or activation for the target emotions. Indeed, in the past it has often been assumed that, contrary to the face, which is capable of communicating qualitative differences between emotions, the voice could only signal levels of physiological arousal (see Juslin & Laukka, 2003; Juslin & Scherer, 2005; Scherer, 1979, 1986, for a detailed discussion). This conclusion is erroneous, as demonstrated by the fact that judges are almost as accurate in inferring different emotions from vocal as from facial expression (see below). Indeed, several studies showed that acoustic properties of speech vary with respect to the emotional quality, intensity, and context (e.g., Bachorowski & Owren, 1995). Furthermore, studies of emotion encoding in voice from a crosscultural perspective (comparing tonal and nontonal languages) found that fundamental frequency and speech rate were used in different ways by speakers of different cultures (Anolli, Wang, Mantovani, & De Toni, 2008; Ross, Edmondson, & Seibert, 1986). This finding confirms that vocal expression does not exclusively signal physiological arousal but also qualitative and culturally sensitive differences between emotions. Qualitative differentiation of emotions in acoustic patterns, apart from arousal, has been difficult to demonstrate because (a) only a limited number of acoustic cues have been studied and (b) arousal differences within emotion families have been neglected (see Scherer, 2003, for further detail). The specification and investigation of additional acoustic parameters with respect to emotion will be essential to our further understanding of vocal expression and perception of emotion, as will the careful consideration of the relationship between arousal and emotion type.

### Vocal decoding studies

Most of the extant research within this modality has been designed analogously to that done with facial expression. In the case of vocal expression, these studies examine the extent to which listeners are able to infer speaker emotions from speech samples. Most of these studies have operationalized emotion as that which is enacted by professional actors, although experienced or induced emotion stimuli have been used in some studies. These stimuli may also differ along the dimension of content: There is no such thing as contentless vocal expression, and researchers have chosen different types of content for their experiments, from meaningful sentences, to sentence fragments, to strings of nonsense syllables (simulated speech), or simply to sustained vowels.

Stimuli specifics notwithstanding, listeners are asked to infer the nature of the portrayed emotions, generally on rating sheets with standard lists of emotion labels, allowing computation of the percentage of stimuli per emotion that were correctly recognized. All the studies in this area have found better-than-chance accuracy in cross-cultural recognition of vocally expressed emotion. Table 2 (vocal expression) reports the average recognition accuracy (based on some of the major studies) of six emotions by Western and non-Western encoders and decoders from vocal expressions.

Vocal emotion expressions are inherently dynamic and generally much less iconic than facial expression patterns (with the exception of affect vocalizations like laughter or vocal emblems—e.g. "Yuk"; Scherer, 1994). It is thus not surprising that the overall recognition accuracy is somewhat lower than for facial expression. However, we can expect to have higher accuracy scores when using only iconic vocal expressions or vocal bursts (Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). Recently, for example, Hawk and colleagues (2009) found that accuracy scores for nonlinguistic affective vocalizations and facial expressions were almost equivalent across nine emotions and both were generally higher than the accuracy for speech-embedded prosody (with a few emotion specific exceptions, e.g., surprise).

In the data reported in Table 2 (vocal expression), the best recognized emotions are sadness and anger. This may reflect the importance of the arousal dimension as discussed earlier, with sadness being very low and anger rather high on that dimension. As there is at least one pertinent study, a comparison of non-Western encoders with Western encoders is possible, in both cases for Western decoders. The accuracy percentages are somewhat lower for non-Western encoders but the differences are not very strong (and even in the opposite direction for sadness). The few studies in which non-Western decoders were asked to recognize emotions portrayed by Western encoders show very low accuracy percentages—particularly for happiness, fear and anger. This is probably due to the fact that in vocal expression there are no such unequivocal cues as the zygomaticus or frontalis action found in facial expression. Rather, as fundamental frequency (pitch) of the voice is a reliable cue for arousal, highly aroused emotions are likely to be confused if the perceiver focuses too much on this cue.

Overall, additional research is obviously necessary in this area, simply to generate a larger body of evidence. Research done on vocal expressions of emotion paired with dynamic facial expressions may be especially enlightening for understanding the specific contribution of each of the two expressive modalities to the emotion inference process in the framework of multimodal communication (see, for example, Collignon et al., 2008).

### Posture/gesture

So far we have focused on the expression of emotion in the face and voice. There are, of course, other channels of expression. Do cultural boundaries play a role there? To date, empirical evidence is sparse on encoding and decoding of bodily cues. It is thus difficult to draw conclusions, but quite easy to identify areas of potentially very fruitful future research.

*Postural/gestural encoding studies.* Research on posture and gesture has been hampered by the lack of a reliable and validated coding scheme comparable to the FACS system for facial actions (see Scherer & Wallbott, 1985; Wallbott, 1985), Using the Giessen system for movement notation (cf. Scherer, Wallbott, & Scherer, 1979), Wallbott and Scherer (1986) coded the body movement behavior of six professional actors (portraying, in dyadic interactions, four emotions—joy, sadness, anger, surprise) into discrete categories such as hand movement illustrators or adaptors or body orientation toward or away from the interaction partner. Results showed a number of specific gesture and movement patterns, especially for sadness.

Using an eclectic coding system based on the earlier work in the group, Wallbott (1998) analyzed a sample of 224 video recordings, in which actors and actresses portrayed the emotions of elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride, and boredom via a scenario approach (see Banse & Scherer, 1996).

Results showed that some emotion-specific movement and posture characteristics seem to exist, but that for body movements differences between emotions can be partly explained by the dimension of activation.

Recently, Dael, Mortillaro, and Scherer (in press-a) have developed the Body Action and Posture (BAP) coding system, for the time-aligned microdescription of body movement on an anatomical level (different articulations of body parts), a form level (direction and orientation of movement), and a functional level (communicative and self-regulatory functions). Applying the system to a corpus of acted emotion portrayals (GEMEP, see below), its comprehensiveness and intercoder reliability—at three levels: (a) occurrence, (b) temporal precision, and (c) segmentation—could be established. It should be noted in passing that body posture is not only a symptom of an individual's emotional state; it can also serve regulatory functions. Thus, Riskind (1984) changed participants' postures in a standard learned helplessness setting. Results indicate that when a slumped posture was "inappropriate" to the current situation (a participant had just succeeded), the slumping seemed to undermine subsequent motivation as well as feelings of control. But when "appropriate" (a participant had experienced failure or helplessness), slumping minimized both feelings of helplessness and depression and motivation deficits.

Extensive cultural influence has been demonstrated in the case of gestures, particularly when the gestures are relatively iconic or emblematic (Morris, 1977). However, from examinations of his film records from various non-Western cultures, Eibl-Eibesfeldt (1989) reported that there are both universal and culturally specific bodily gestures. For example, foot stamping as a sign of anger appears to be universal, whereas presentation and squeezing of a breast as a signal of fear is found in only a few of the societies studied.

However, this is not necessarily indicative of a culture by affect interaction (in terms of determining gestural expression output). It may be the case, for example, that gestures are to a lesser extent determined by emotion, as has been suggested by Rimé and Schiaratura (1991). Cultural influence could be realized through other routes, including differential modes of dress (ultimately linked to location in differential climates), for example, which make gestures with different body parts more or less salient.

Future work examining possible cultural "dialects" in the production or perception of bodily cues could potentially make progress by starting to systematically study gestures (see below). Dialect theory may be particularly appropriate for gesture, which is thought to be relatively stereotyped and symbolic (and thus relatively similar to language).

*Postural/gestural decoding studies*. In terms of experimental examinations, some extant work has been done on affect and posture generally, but very little with an eye toward universality versus cultural specificity. There is almost none on gesture, despite the expectation of greater cultural patterning within this domain. General inferences can be drawn, however, which may be useful for guiding future research that does focus on culture and its influence on all bodily cues. Some inferences can be made, as well, by comparing culturally nonspecific research on bodily cues with similar, culturally specific research within other expression domains.

Research on the effect of postural variables, such as orientation (toward an interaction partner), degree and direction of trunk lean, head orientation, shoulder orientation, leg orientation, arm openness, and leg openness, or on affect type and intensity (Argyle & Kendon 1967; Ekman, 1965; Exline & Winters, 1965; Kendon, 1967; Mehrabian & Friar, 1969) has suggested that these variables may have utility in terms of emotion expression, but this may hold only for gross categorizations (e.g., positive vs. negative). This may be an artifact of the relative number of potential feature configurations or of experimental trends (most of these studies have looked only at gross affect categories, with no clear a priori reason for so doing), but it has also been suggested that there is specific utility in channels for gross expression, distinct from that of channels that may be more precise (e.g., the voice) (Ekman, 1965; Ekman & Friesen, 1967). If there is differential utility, perhaps it varies by culture, or simple ingroup and outgroup status. Notably, expressive utility is also limited for gesture because of relatively little variation in form (not function, as is suggested in the case of posture). Differences in form and function between gesture and posture are an additional area in which cultural influences may be brought to bear.

The use of similar experimental paradigms in the examination of bodily and other expression types may also allow for the assertion of some basic inferences as to potential cultural specificity; specifically for posture, there are dimensions along which the existing research can be compared with that in other modalities. For example, one study looked at the influence of proxemic variables on inferred attitudes (by a perceiver, as opposed to how these variables relate to emoter-reported affect), a paradigm that is very similar to the one used to investigate potential cultural boundaries within the communication of affect via vocal and facial cues (Mehrabian, 1968b). Furthermore, most of the literature on culture posits a multidimensional structure of nonverbal behavior (Kudoh & Matsumoto, 1985; Mehrabian, 1968a, 1972, 2007; Mehrabian & Friar, 1969), which bears significant resemblance to the dimensional structure reported for facial and vocal expression data; each of the postural variables measured therein are proposed to speak to one of these dimensions (see Osgood, 1966; Schlosberg, 1954; Williams & Sundene, 1965).

With regard to inferences about universality versus specificity, only very general conclusions can be suggested. We can expect that we will see correlations between posture and emotion that hold crossculturally in the cases in which there are "third variables" that engender independently both certain emotions and certain physiological states (e.g., in the context of depression, which may make more likely both sadness and lax muscle tension). In terms of emotion-determined postures and expressions, there is some support for both universality and specificity (Morris, 1977), but care must be taken when making inferences from this point because hard evidence is still very limited. Eibl-Eibesfeldt's multicultural film library (http://erl.orn.mpg.de/~fshuman/en/eindex.html) also allows for some crosscultural comparison. Finally, one of the postural studies cited earlier—which utilizes dimensions similar to those suggested by appraisal theorists—was carried out with Japanese participants, allowing for some comparisons to be made (as all others were done with participants of Western culture); no significant postural differences by culture are evident (Kudoh & Matsumoto, 1985).

If it were the case that emotion was a significantly less important determinant for gesture, as compared with other forms of emotion expression, we would have to analyze gestural evidence in a unique fashion. A multicultural inventory of expressions would still be highly desirable, but when considering individual expressions, we would have relatively less confidence that any difference, emotion held constant, was due to cultural influence until we were also able to identify, and hold constant, any other influential factor. Identifying whether other factors play an important role, and what those factors may be if they exist, is thus a priority on this front.

### *Multimodal expressions*

As mentioned at the outset, most work in this area has almost exclusively focused on individual expression modalities—primarily the face, and to a somewhat lesser extent the voice, and only very rarely gesture and posture. Recently, however, several efforts have been made to look at multimodal expressions, even though, also in this case, most of the studies focused on emotion decoding. Multimodal expression refers to synchronized vocal, facial, and gestural expression patterns in a particular emotional expression episode.

*Multimodal encoding studies.* There is an alarming shortage of studies investigating how emotions are encoded in different expressive modalities when more than one modality is used at the same time. Real everyday communications engage multiple expressive modalities simultaneously. Unfortunately, we know practically nothing about how expressers use (or decide not to use) multiple modalities in a more or less coherent fashion to communicate or encode an emotion state or appraisal. Most researchers have focused on single modalities because of their special research competences in the respective modality but also because it is extremely difficult to study multimodal encoding of emotion. The subject is complex from both a theoretical (e.g., is there a prototypical multimodal configuration for some emotions? Is there a unique efferent mechanism that is shared between different modalities? How should synchronization be defined?) and an empirical point of view (e.g., how to obtain multimodal expressions that are valid, reliable, and of good quality to allow dynamic microcoding of behavior; how to analyze dynamic behavior and operationalize synchronization).

There have been efforts to analyze audiovisual recordings of emotional expressions on the fly (from the media, for example), in public settings (e.g., emotions upon baggage loss in the airport; Scherer & Ceschi, 2000), or by inducing low-intensity emotions in the laboratory (see Cowie et al., 2010; Douglas-Cowie, Campbell, Cowie, & Roach, 2003). Although these corpora have been compiled and annotated, no systematic comparison of the expressions that they contain is available. Furthermore, there are serious concerns about the reliability of the emotion labels that are attributed to these public expressions as well as about their validity as "natural" emotion expressions.

The need for reliable material of high quality and the underdeveloped research on this topic require using expression corpora that have been obtained in a systematic way. There are now a few corpora in which enacted expressions have been systematically recorded. In particular, Scherer and his group have been using enacting procedures (Stanislavski or method acting approaches, as well as verbal imagery) to obtain integrated expressions in the face, voice, and body. The first example is reported in the early study by Wallbott and Scherer (1986) cited above. In this study, the authors objectively analyzed both gesture/body movement and vocal cues produced by six professional actors who enacted different emotion interaction scenarios. Results showed that the four emotions were expressed differently, in terms of both gestures and vocal characteristics, and behaviors of both modalities had a significant effect on decoding accuracy. In a subsequent corpus, the "Munich" corpus (Banse and Scherer, 1996), 12 professional stage actors (six men) portrayed 14 different emotions—the anger family (hot anger, cold anger), the fear family (panic fear, anxiety), the sadness family (despair, sadness), the happiness family (elation, happiness), and interest, boredom, shame, pride, disgust, and contempt—that were recorded on high-quality videotape. Stimuli were recorded multimodally, and were first analyzed in each expressive modality separately. The acoustic analyses of the vocal expression (audio) are reported in Banse and Scherer (1996); the objective analysis of facial expression (using FACS) in Scherer and Ellgring (2007a). Subsequently, Scherer and Ellgring (2007b) reported a first analysis of multimodal expressive behavior configurations. The authors considered acoustic features, FACS codes, and body movements that were used for encoding the 14 emotions and were able to identify three clusters of multimodal behavior: agitation, resignation, and joyful surprise. These clusters grouped behaviors pertaining to different modalities and were used to portray different emotions, suggesting that patterns of multimodal behavior are not emotion-specific.

A new, dynamic, multimodal corpus of emotion expressions, the Geneva Multimodal Emotion Portrayals (GEMEP; Bänziger & Scherer, 2010), produced digital audiovisual records of 10 professional French-speaking theater actors (five men) enacting 18 emotions. Twelve "core emotions" had been chosen in such a way as to represent the quadrants of a valence by arousal design, allowing comparison of the respective emotions in terms of the dimension differences: positive/high arousal—pride, joy; amusement; positive/low arousal—interest, pleasure, relief; negative/high arousal—hot anger, panic fear, despair; negative/low arousal—irritation, anxiety, sadness. Additionally, the following emotions were enacted by five actors respectively: admiration, tenderness, disgust, contempt, surprise, relief, and interest. Both standard nonsense sentences and vocalizations (schwa sounds) were used. Actors also produced portrayals with different intensities and masked portrayals. The original set of 1260 expressions of the GEMEP corpus was validated using a series of rating studies and reduced to a smaller subset of well-recognized stimuli—*GEMEP Core Set* (see "Multimodal decoding studies" below). The multimodal stimuli of the core set have been dynamically analyzed in terms of facial expressions, acoustic characteristics, and body movements. The first results on the facial expression productions, using dynamic frame-by-frame FACS coding (i.e., considering the temporal phases of each action unit: onset, apex, and offset), have appeared in Mortillaro, Mehu, and Scherer (2011). These authors showed that, contrary to the assumption that all positive emotions share the smile as a common signal but lack specific facial configurations, the frequency and duration of several action units differ between the emotions of interest, pride, pleasure, and joy, indicating that actors do not use the same pattern of expression to encode them. Furthermore, authors suggested that these differences can be plausibly interpreted by adopting an appraisal perspective. Mehu, Mortillaro, and Scherer (2011) investigated the impact of two subsets of facial action units (reliable AUs and versatile AUs; see Ekman, 2003) on the accuracy of identification of the emotion conveyed and its perceived authenticity. Activity of the reliable AUs had a stronger impact than that of versatile AUs on accuracy of labeling and perceived authenticity of emotional portrayals. A paper reporting a comprehensive analysis of the dynamic facial behavior used to encode the 18 emotions and its relationship with emotion appraisals and action tendencies is currently in preparation.

Dael, Mortillaro, and Scherer (in press-b) analyzed body movements, gestures and, postures that were used to encode the different emotions. Authors applied the BAP coding system (Dael et al., in press-a) to the GEMEP stimuli and found that several patterns of body movement systematically occur in portrayals of specific emotions, allowing a gross emotion differentiation. While a few emotions were prototypically encoded by one particular pattern, most were variably expressed by multiple patterns, many of which can be explained as reflecting functional components of emotion such as modes of appraisal and action readiness (Dael et al., 2011b).

The GEMEP stimuli were also used to analyze how emotions were encoded vocally. Actors encoded emotions using either nonsense utterances or vocalizations. The results of the acoustic analyses of the nonsense utterances (Goudbeek and Scherer, 2010) replicated results reported in literature (e.g., Banse & Scherer, 1996) and determined the relative contribution of the respective vocal parameters to the emotional dimensions arousal, valence, and potency/control. The results show that although arousal dominates for many vocal parameters, it is possible to identify parameters, in particular spectral balance and spectral noise, that are specifically related to valence and potency/control. The analysis of the affect bursts (vocalizations) showed significant emotion main effects for 11 of 12 acoustic parameters reflecting three major factors: ''tension,'' ''perturbation,'' and ''voicing frequency'' (Patel, Scherer, Sundberg, & Björkner, 2011). Another study focused on the production mechanisms (estimating subglottal pressure, transglottal airflow waveform, and vocal fold vibration) and showed that each emotion appeared to possess a specific combination of acoustic parameters reflecting a specific mixture of physiologic voice control parameters (Sundberg, Patel, Björkner, & Scherer, 2011).

The corpus represents one of the few examples of genuinely multimodal emotion expressions, i.e., emotion expressions that were not controlled or constrained in any modality. Although analyses until now focused on individual expressive modalities, the GEMEP expressions provide an appropriate dataset for studying emotion encoding in a multimodal perspective. The key element of multimodality is the synchronized unfolding of the behavior in the different modalities. Therefore, multimodal analysis requires dynamic stimuli and temporally aligned microcoding of behaviors. Currently, several researchers are trying to develop psychologically valid multimodal material, but this material seems more suitable for decoding studies than for encoding studies. For example, Hawk and colleagues (2009) obtained multimodal recordings of emotion expressions, but their main goal was to develop sets of reliable and valid stimuli to be used in decoding studies. The authors derived three subsets of stimuli (two sets of vocal expressions—affect vocalizations and speech—and one of silent facial expressions) from three separate recordings. In each of these recordings, expressers were invited to focus on expressing the respective emotion in one modality (even though both the facial and vocal modalities were recorded in each case). Therefore these recordings provide reliable and validated stimuli for multimodal decoding studies, but are not ideal for studies on how emotions are encoded multimodally.

*Multimodal decoding studies.* Research on multimodal perception of emotion expressions is slightly more advanced than the work on emotion encoding. We can identify three lines of research pursued in this domain. A first line of research investigated emotion perception and communication between mothers and children, a form of communication in which synchronization between expressive modalities seems to have a key role (Gogate, Bahrick, & Watson, 2000). Infants of a few months of age are already sensitive to multimodal information (for a review, see Walker-Andrews, 1997), and they use this information for acquiring speech (Legerstee, 1990) and to discriminate emotional expressions (Caron, Caron, & MacLean, 1988).

The second line of research includes those studies that investigated the relative contribution of two channels to emotion perception when the cues pertaining to different channels provide somewhat conflicting information. For this kind of research, scholars generally use bimodal stimuli artificially created in the laboratory by pairing two unimodal stimuli. Studies involved both children—with the goal of determining which modality is dominant at which age (for example, Robinson & Sloutsky, 2004; Shackman & Pollak, 2005)—and adults. Massaro and Egan (1996), for example, combined facial and vocal cues to create congruent and incongruent expressions. The authors found that observers integrated information from both sources to form their judgment, and when one channel was ambiguous the other became more influential. Similarly, studies in which stimuli combined faces and postures found that body influences the judgments on the emotion expressed in the face (Meeren, van Heijnsbergen, & de Gelder, 2005). Considered together, these studies speak in favor of the idea that emotion perception is inherently multimodal. However, the stimuli that were used are not real multimodal expressions and the results are not informative of how information is actually integrated by the observers and what the exact contribution of each modality is—the implicit assumption of these studies is that cues of different modalities can be manipulated separately and added to form a multimodal expression. Conversely, signals belonging to different modalities can be perceptually combined in many different ways, and the inferences based on a multimodal signal may be different from the sum

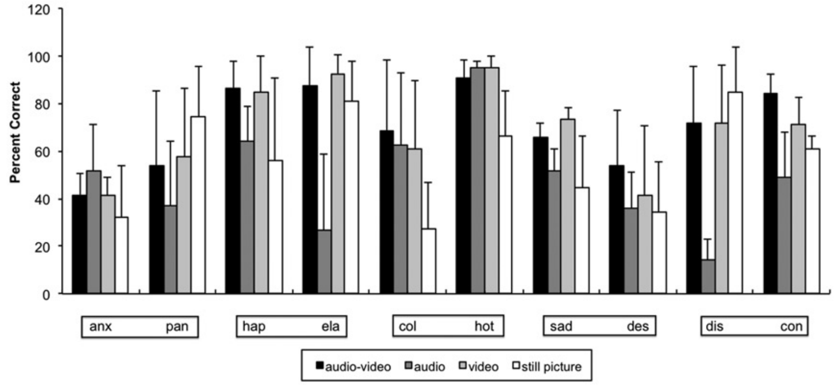


**Figure 2.** MERT corpus: Mean accuracy scores (percentages) for five families of emotions presented in audio–video, audio, video, and still picture modalities. Anx = anxiety; pan = panic fear; hap = happiness; ela = elation; col = cold anger; hot = hot anger; sad = sadness; des = despair; dis = disgust; con = contempt (reprinted from Bänziger et al., 2009, p. 697).

of the inferences based on the unimodal signals considered separately (Partan & Marler, 1999)

A third line of research is directed at studying real multimodal stimuli, for which emotional content is known and the behavior is annotated. In these studies, scholars pair the investigation of the decoding process with the analysis of the encoding process; therefore, the stimuli are not created in the lab or determined on the basis of fixed instructions given to the posers, but rather freely produced by expressers based on induction or enacting techniques. In the decoding part of these studies, researchers normally show the stimuli in different perceptual conditions. So for example, Wallbott and Scherer (1986) showed the same stimuli using four modes of presentation (audio–video, video only, audio only, filtered audio), to different groups of naive judges. The results indicated that differences between channels and between actors strongly affect decoding accuracy. Specifically, overemphasis of behavioral cues characteristic for certain emotions resulted in reduced decoding accuracy.

More recently, Bänziger et al. (2009) reported the recognition accuracy data for the Munich corpus described earlier, which were obtained as part of the development of the Multimodal Emotion Recognition Test (MERT). Figure 2 shows the average percentages of correct answers for the emotions across four presentation modalities (audio only, video only, audio–video, or still picture stimuli) obtained in this study. The data reveal statistically significant interactions (see Bänziger et al., 2009, p. 697) between presentation modality and emotions. Most noticeably, for disgust and elation, portrayals presented in audio mode (vocal portrayals) are less well recognized than those including visual (facial) cues. The low accuracy for disgust recognition generalizes across studies on vocal expression (see Table 2)—the reason might be that disgust is often a brief, burst-like emotion that does not affect lengthier utterances (see also Banse & Scherer, 1996; Hawk et al., 2009). The low accuracy for elation is explained by high pitch, which is often seen as high arousal and confused with anger and fear. Video clips (dynamic facial stimuli) appear to facilitate accuracy in identification in comparison to still pictures (static facial stimuli), with the noticeable exceptions of expressions of panic and disgust. The reason why this is different in direction from the results reported in Table 2 (facial expression) can probably be sought in the fact that the still photos were extracted from the dynamic video and thus were not carefully constructed as is the case in facial expression photo corpora generally. Figure 2 also allows us to examine the differences in recognition accuracy between the two members of an emotion family. Generally, the more active or intense member is recognized with higher accuracy, although there are exceptions. Thus some of the more intense emotions are more frequently confused in the audio condition, probably because of the fact that pitch and amplitude cues rise with higher arousal for several emotions. Furthermore, despair as the presumably more intense member of the sadness family is more frequently confused in all modalities, probably because it tends to be a mixed emotion with strong elements of anxiety and fear in addition to resignation. It should be noted that the results from this dynamic multimodal corpus do not confirm, as illustrated in Figure 2, that recognition accuracy is lower when dynamic expressions are used rather than still pictures. Here, with two exceptions (panic fear and disgust), dynamic video

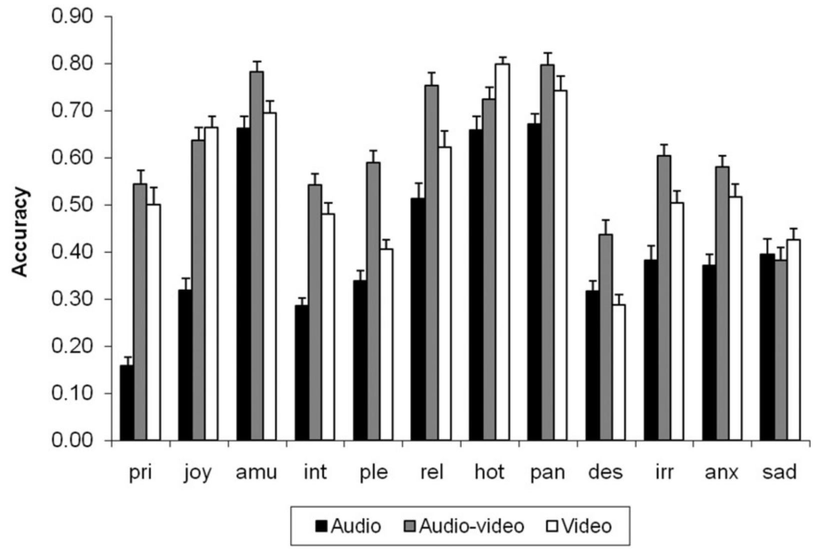


**Figure 3.** GEMEP corpus: Mean accuracy scores (proportions) for 12 emotions varying in valence and arousal presented in audio–video, audio, and video modalities. Positive/high arousal—pri = pride; joy = joy; amu = amusement; positive/low arousal—int =interest; ple = pleasure; rel = relief; negative/ high arousal—hot = hot anger; pan = panic fear; des = despair; negative/ low arousal—irr = irritation; anx = anxiety; sad = sadness (reprinted from Bänziger & Scherer, 2010, p. 282 by permission of Oxford University Press).

and audio–video presentations tend to be recognized at least as well and often more accurately.

Figure 2 shows the accuracy percentages for the individual emotions. However, it should be noted that these accuracies and particularly the confusions or errors require a more sophisticated analysis in this case, as a confusion between family members is obviously much less of an error that a confusion with a completely different emotion from another family. This is why it has been suggested to compute, in addition to individual recognition indices, so-called ''family recognition indices'' that measure the ability to perform a rough identification of family membership (see Bänziger et al., 2009, p. 696).

The GEMEP corpus has also been used in emotion multimodal decoding studies. In a first comprehensive decoding study, 1260 emotion expressions were presented to naïve judges in three different presentation modes: video-only, audio-only, and audio–video. The results (see Figure 3) again showed a significant interaction effect between emotions and modality (Bänziger & Scherer, 2010, p. 280). Again, accuracy scores were highest in the conditions in which visual (facial) information was available and lowest in the audio-only condition, confirming previous unimodal studies—as reported in Table 2. However, this lower accuracy in the audio-only mode may be due to the absence of clear prototypical cues in the voice, at least for some emotions, as compared to the face, where there a number of very clear discrete cues, such as the smile for positive emotions. With a notable exception (hot anger), audiovisual presentations are judged more accurately than video only, presumably because of the combination of relevant cues stemming from two different modalities. However, the differences are relatively small, suggesting that there is a high degree of redundancy. Comparing the valence and arousal dimensions underlying the design of the emotion selection, again the more aroused emotions are generally recognized more accurately. However, here also we find some exceptions, for example amusement and relief. This could be due to laughter and sighs used by some of the actors in these expressions. Again, despair, although considered a high arousal emotion, is less well recognized, probably because of the elements of anxiety or fear present in this mixed emotion.

Based on the results of this first decoding study, Bänziger et al. (2011) selected a subset of expressions that had been judged in the earlier study as typical and believable examples of the target emotions.[6] A second decoding study using this subset of expressions confirmed previous findings, that is the stimuli are most accurately decoded in the audio–video mode and least accurately in the audio-only mode. In addition, several

[6] Judges who participated in the decoding study also rated each expression for its degree of believability, defined as ''the capacity of the actor to communicate a natural emotional impression'' (Bänziger & Scherer, 2010, p. 278).

modality–emotion interactions attained significance (see Bänziger et al., 2011).

Table 2 (multimodal expression) displays the mean results from the multimodal studies of emotion expression that have been published thus far (to our knowledge). It should be noted that in two of the three studies synthesized here, emotion families have been used with the need, as described above, to evaluate confusion patterns and accuracy differently. For this reason, the accuracy coefficients shown in Table A4 in the Appendix and in Table 2 (multimodal expression) consist of average coefficients (computed as the sum of the accuracy percentages for each emotion member plus the respective confusion with the other family member and then obtaining the average of the two values).

The fact that additional data need to be collected is especially obvious here. First of all, as all three studies stem from our own laboratory, it is imperative to obtain comparative results from other laboratories. Furthermore, at the time of this writing, we were unable to find a single study examining either the encoding or the decoding of multimodal expressions by members of any non-Western culture. Another reason for encouraging more work with multimodal corpora is the probability that the expressions are less prototypical and achieve higher ecological validity through the enacting procedures used, which require the synchronized use of all expression modalities (see Scherer & Bänziger, 2010). As a consequence, the results obtained exclusively with still photo corpora (but maybe also of pure audio and pure video corpora) need to be treated with some caution, and it seems advisable to turn increasingly toward research that uses dynamic multimodal corpora.

## Accuracy and error: Confusion matrices

At this point, it is imperative to point out that the recognition accuracy (often expressed as a percentage) is not the only, and may not be the best, indicator as to the comparability of emotion inferences from expression across cultures. In fact, the accuracy percentage indicator has been often accused of biases. For example, if judges have a tendency to check certain emotions more frequently, they will be more precise on these emotions just by chance (Wagner, 1993, has suggested a correction procedure for this bias). Furthermore, if there are only a few emotions (e.g., only about five or six) and if there is only one positive emotion, which is often the case, the process of recognition is likely to become one of simple discrimination based on highly salient cues (such as the smile) or on frequency of occurrence and guessing. Russell (1994) has published a strong critique of the methods employed in this research area, namely that the forced-choice paradigm may have artificially forced agreement, a critique that has been equally strongly rebutted by Ekman (1994). The positive effect of this debate has been that it is now considered state of the art to provide judges with the option of "none of these," which minimizes the pressure to guess (Frank & Stennett, 2001).

A powerful check on potential biases in the judgment procedure is provided by the computation of a confusion matrix. The off-diagonal of the matrix indicates the extent to which the patterns of errors in identification of emotion expressions are comparable across decoder groups, providing interesting information about the underlying cue utilization patterns. Table 4 shows an example of a confusion matrix, compiled from a selection of studies examining the identification of emotion expressed in still photographs, separately for Western and non-Western decoders; similar comparisons but across emotion expression modalities (e.g. facial static, facial dynamic, vocal) instead of decoder culture can be found in the literature (see Bänziger & Scherer, 2010). It is essential to point out that, whereas the recognition accuracy in the non-Western decoders studied is lower, the pattern of errors made is highly comparable. There is one exception for the emotion of contempt and it may not be an accident that the universality of the expression of exactly this emotion has been debated (Ricci-Bitti, Brighetti, Garotti, & Boggi-Cavallo, 1989; Rosenberg & Ekman, 1995).

This issue is of great importance, especially in research in which only a few emotions are studied (and often only one positive emotion), as it permits one to differentiate between emotion differentiation (for example excluding unlikely alternatives from a small list) and emotion recognition (using a specific prototypical pattern of cues to identify a particular emotion). Furthermore, if most of the confusions are among members of the same emotion family, this is obviously much less problematic for the assumption that there are valid expressive cues than if fundamentally different emotions are confused (see Bänziger et al., 2011, for a more detailed discussion and a utilization of the confusion patterns in differentiating test results). This is why it has been suggested above to use family recognition indices, including intrafamily confusions, in cases where a large number of emotions organized by families are studied.

**TABLE 4**

Confusion matrix: Predicted and participant-chosen emotion labels for facial expressions (still photographs)

| *Participant chosen label* | *Predicted label* | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Anger* | | *Contempt* | | *Disgust* | | *Fear* | | *Happiness* | | *Sadness* | | *Surprise* | | *Other* | |
| *Decoders* | W | NW | W | NW | W | NW | W | NW | W | NW | W | NW | W | NW | W | NW |
| Anger | 57 | 44 | 7 | 8 | 11 | 16 | 9 | 11 | 0 | 1 | 4 | 5 | 1 | 2 | 4 | 6 |
| Contempt | 12 | 11 | 39 | 30 | 17 | 17 | 2 | 2 | 0 | 1 | 9 | 4 | 1 | 1 | 9 | 6 |
| Disgust | 8 | 9 | 20 | 7 | 57 | 42 | 6 | 8 | 0 | 1 | 4 | 2 | 2 | 2 | 5 | 4 |
| Fear | 4 | 4 | 1 | 2 | 2 | 5 | 50 | 31 | 0 | 1 | 2 | 5 | 8 | 7 | 3 | 5 |
| Happiness | 0 | 1 | 1 | 4 | 0 | 1 | 1 | 2 | 84 | 80 | 0 | 2 | 17 | 16 | 9 | 9 |
| Sadness | 3 | 4 | 5 | 5 | 4 | 5 | 1 | 3 | 0 | 1 | 59 | 54 | 1 | 2 | 8 | 10 |
| Surprise | 8 | 9 | 4 | 6 | 1 | 3 | 28 | 35 | 5 | 3 | 2 | 3 | 64 | 54 | 5 | 5 |
| Other | 9 | 19 | 24 | 40 | 9 | 13 | 5 | 13 | 10 | 13 | 20 | 25 | 8 | 16 | 57 | 57 |

Predicted and participant-chosen emotion labels for facial expressions (portrayed in still photographs), as reported in two crosscultural studies of emotion expression (Elfenbein et al., 2007—Canadian and Gabonese samples; Yik & Russell, 1999—Canadian and Chinese samples). Figures given are the percentage of trials in which participants chose the specified label for each target. Columns are the emotions that the targets were instructed to express; rows are the labels participants were given, from which they were to choose *one*, according to the emotion they believed the targets to be experiencing. For each target emotion, the nonshaded column on the left (W) displays the rate at which decoders from a Western culture chose each label, and the shaded column on the right (NW) displays the rate at which decoders from a non-Western culture chose each label (all encoders were Caucasian, and Canadian or American). The heavily shaded areas correspond to correct answers. Total percentages may deviate from 100%, as reported rates were rounded.

## CULTURAL INFLUENCES

### Encoding

#### *Display rules*

Early evidence showed culturally specific expression production when the expression was produced in a social context, but largely universal expression production when the expresser thought he or she was alone (Ekman & Friesen, 1969). The authors held this work to reconcile the earlier findings, which they claimed to support universality, with the findings of anthropologists, which reported dissimilar facial expressions across cultures. Ekman and Friesen (1969) specifically termed this mechanism ‘‘display rules.’’ Complementary evidence of decoding rules (see ‘‘Decoding’’ below) soon followed (Matsumoto & Ekman, 1989).

Except for questionnaire studies, there have been few empirical encoding studies aimed at proving or disproving the existence of such rules (Gaspar, 2006; Malatesta & Haviland, 1982; Matsumoto, 1990, 1993; Wagner, 1990; for other supporting evidence, see Beaupré & Hess, 2005; Matsumoto, Kasri, & Kooken, 1999; Matsumoto & Kudoh, 1993), but encoding and decoding rules are widely cited within the literature as factors that must be taken into account from the outset (see for example Cohen, Sebe, Garg, Chen, & Huang, 2003; Krauss, Apple, Morency, Wenzel, & Winton, 1981; Rinn, 1984) and/or as factors which may explain results that deviated from what the authors expected (see for example Kleinsmith, De Silva, & Bianchi-Berthouze, 2006; Schimmack, 1996; Wagner, Lewis, Ramsay, & Krediet, 1992; Wagner, MacDonald, & Manstead, 1986; Winkelmayer, Exline, Gottheil, & Paredes, 1978). The existence of encoding and decoding rules has furthermore been integrated into newer theories of emotion expression that have to account for the findings of at least some cultural specificity (see for example Elfenbein & Ambady, 2002; Elfenbein, Lévesque, Beaupré, & Hess, 2007; Mandal & Ambady, 2004).

#### *‘‘Push’’ and ‘‘pull’’ factors*

Scherer and collaborators have suggested to use the term ‘‘push effects’’ to denote the core features of an expression that are pushed to the surface by the operation of physiological changes in the service of adaptation, and ‘‘pull effects’’ to refer to the effects of culture and social context (Kappas et al., 1991; Scherer, 1985, 2003). These effects specify the ideal form the expression could take, given the particular social context and control needs, including the person’s strategic intentions. In order to demonstrate the nature of the interaction between push and pull effects that account for particular patterns of expression, research has to be specific in identifying cultural norms or expectations, and individuals’ intentions.

Obviously, both of these factors are operative at all times and it is exceedingly difficult to find cases in which only one of the two factors will dominate (except, possibly, in the case of a purely tactical use of emotion expression). For this reason, the relative importance of push and pull factors is very difficult to establish empirically.

We should aim to understand how any cultural variation in emotion expression that we might be able to attribute to differences in encoding according to culture is the result of one or the other (or both) of these processes. For example, we noted earlier that within the context of facial expressions, there appear to be learning effects, such that perceivers of one culture can improve their accuracy of identification of expressions encoded by members of another culture by becoming more familiar with that culture. This would suggest that, to some degree, cultural variation within the expression process is realized in the encoding process and is due to learned pull factors.

### Transmission

This factor relates to the route by which the distal cues are transmitted to the observer, i.e., via light waves for visual cues and acoustic waves for auditory cues. This difference is of course essential in environments in which members of a group cannot see each other or in which noise masks auditory signals. It can be expected that evolutionary factors have played a key role in shaping the emotional signaling system on the basis of such constraints.

Context is another example of an outside influence that affects transmission and that must be accounted for when making any inference about the emotion expression process. It is well known that context plays a central role in communication and perceptual processes. With regard to emotion expression specifically, context could interact with both the expression and the perception processes. In the case of expression, context may help determine pull effects. Specifically, there is evidence that affect type (positive vs. negative) influences the degree to which expressers will display the effects of pull factors. A positive affect state appears to be associated with less culturally normative expressive behavior (i.e., that which might result from a lesser reliance on pull factors), and a negative affect state with greater bounding to cultural norms (Ashton-James, Maddux, Galinsky, & Chartrand, 2009). In the case of perception, context could affect both how the distal cues determine proximal percepts and how attributions of emotion state are made on the basis of those percepts. Indeed, this has already been shown to be the case: Students with a lower-class background performed better than those with an upper-class background in recognizing emotional expressions, presumably because of the importance of interpersonal skills in their daily lives (Kraus, Côté, & Keltner, 2010). Another interesting type of context effect has been shown by Masuda et al. (2008) and Leu et al. (2010), who tested the hypothesis that in judging people's emotions from their facial expressions, Japanese, more than Westerners, incorporate information from the social context. Their participants viewed cartoons depicting a happy, sad, angry, or neutral person surrounded by other people expressing the same emotion as the central person or a different one. The surrounding people's emotions influenced Japanese but not Westerners' perceptions of the central person. These differences reflect differences in attention, as indicated by eye-tracking data: Japanese looked at the surrounding people more than did Westerners.

### Decoding

It could be the case that certain expressions are less frequently utilized in some cultures, but are nonetheless recognizable to the individuals therein, as would be the case if expression production and perception were innate abilities. There may also be culturally specific conventions for perception and/or inference (see Mesquita et al., 1997). If members of different cultures tend to explain events in different ways, then it is reasonable to expect that they may infer different emotion states from the same expressed cues and/or the same cue percepts. Such a difference in explanatory styles would be an example of differential encoding and/or decoding rules, as discussed above in the ‘‘Encoding’’ section (see also Beaupré & Hess, 2005; Buck, 1984; Elfenbein & Ambady, 2002, 2003; Kleinsmith, De Silva & Bianchi-Berthouze, 2006; Matsumoto, 1989; Matsumoto & Ekman, 1989; Morris & Peng, 1994; Scherer, Banse, & Wallbott, 2001; Thompson & Balkwill, 2006).

### Dialect theory

The data derived from studies utilizing decoding paradigms (see Table 2 for a summary) have been consistently interpreted as evidence for the universality of emotion expression. Given this degree of convergence, this conclusion now appears in textbooks. However, there is currently a great deal of work on recognition accuracy, for both facial and vocal expression, which seems to show an ingroup advantage. A 2002 meta-analysis by Elfenbein and Ambady corroborated individual findings that whereas "emotions [are] universally recognized at better-than-chance levels ... Accuracy was higher when emotions were both expressed and recognized by members of the same national, ethnic, or regional group, suggesting an in-group advantage" (p. 203). The preponderance of subsequent studies has supported this concept (Biehl et al., 1997; Elfenbein & Ambady, 2003; Elfenbein & Mandal, 2004; Markham & Wang, 1996; Scherer et al., 2001). Furthermore, some findings suggest that judgments of intensity (as opposed to judgments of differential emotion) may differ by culture as well. This effect appears to be specific to facial expressions, however (Ekman et al., 1987; Matsumoto & Ekman, 1989; Matsumoto & Kudoh, 1993; Saha, 1973). Further research into why this occurs may yield interesting conclusions as to potential differential utility of the facial and vocal channels of expression, if the specificity of effect is, in fact, a stable finding.

The general universality but notable ingroup advantage effect has been extended and shaped into a *dialect theory* of emotion expression. According to dialect theory, facial and vocal expression can be thought of as having signal variations by culture that are subtle, but sufficiently significant to decrease recognition accuracy by outgroup members (Dailey et al., 2010; Elfenbein et al., 2007). Research that reports the presence of "accents" in facial expression—i.e., the existence of subtle encoding differences by culture even when muscle movements are standardized, such that the cultural identity of the emoter can be identified—further supports such a theory (Marsh, Elfenbein, & Ambady, 2003). Independent studies also support the idea that cultural differences in recognition rates are realized in the process of encoding (rather than decoding) (Elfenbein & Mandal, 2004). However, most studies in this area are not sufficiently well controlled on the encoding side to clearly identify whether it is encoding or decoding differences between cultures (or an interaction of both) that is responsible for the effect and the underlying mechanisms.

Future studies should look specifically at potential dialects in decoding. Although the term *dialect* does not map nearly as well onto these processes in this phase, there is reason to believe that culture may shape the expressions to which we are attuned. An effort should also be made to determine the degree to which groups must differ in order for these dialects to be realized. It may be less than we think; a recent study (Young & Hugenberg, 2010) showed that mere social categorization, with a minimal-group paradigm, can create an ingroup emotion identification advantage even when the culture of the target and perceiver is held constant. This suggests the possibility that culture may not be causally linked to differential encoding and/or decoding, but is somehow otherwise related to something that is. The link between patterns found in the literature and culture may be spurious.

It is evident that the issue of cultural effects on the expression-perception process can only be appropriately addressed if encoding and decoding differences are jointly studied. Two recent studies focusing on vocal expression get closer to the kind of research design needed to disentangle the factors described earlier. In one of these studies, Sauter, Eisner, Ekman, and Scott (2010) examined the recognition of nonverbal emotional vocalizations, such as screams and laughs, across two widely different cultural groups (Western participants and participants from isolated Namibian villages). Vocalizations communicating the so-called basic emotions (anger, disgust, fear, joy, sadness, and surprise) were bidirectionally recognized. In contrast, a set of additional emotions was recognized within, but not across, cultural boundaries. The authors suggest that a number of primarily negative emotions have vocalizations that can be recognized across cultures, whereas most positive emotions are communicated with culture-specific signals. In the other study, Pell, Paulmann, Dara, Allasseri, and Kotz (2009) elicited vocal expressions of six emotions (anger, disgust, fear, sadness, happiness, pleasant surprise) and neutral expressions from four native speakers of four different languages using pseudo-utterances ("nonsense speech") that resembled their native language to express each emotion type. The recordings were judged for their perceived emotional meaning by a group of native listeners in each language condition. Emotion recognition and acoustic patterns were analyzed within and across languages. Although overall recognition rates varied by language, all emotions could be recognized from vocal cues in each language at levels exceeding chance. Anger, sadness, and fear tended to be recognized most accurately irrespective of language. Acoustic and discriminant function analyses highlighted the importance of speaker fundamental frequency (i.e., relative pitch level and variability) for signaling vocal emotions in all languages. The data suggest that although emotional communication is governed by display rules and other social variables, vocal expressions of "basic" emotion in speech exhibit modal tendencies in their acoustic and perceptual attributes that are largely unaffected by language or linguistic similarity. Overall, these recent data suggest that basic or modal emotions may well be encoded and decoded in a rather universal fashion, suggesting some degree of evolutionary continuity. In contrast, cultural context may have a stronger effect on the expression of more subtle emotions that are less tied to fundamental contingencies of human life and more imbued with cultural values and conventions.

## CONCLUSION: NEED FOR ADDITIONAL DATA AND THEORETICAL UNDERPINNING

In this review of the current state of the field on expression research, we have focused on the issue of universality versus cultural specificity and we have attempted to identify most major research efforts that have been published to date. Given the large number of studies in this area, published in different languages and in sources that are not always easy to locate, we may well have overlooked pertinent research. Due to lack of space, we have specifically excluded neuroscience research in this area, where (mostly facial) expression of emotion has become a very popular topic of research (see Adolphs, 2002). Based on the results of the review it seems evident to us that there is a convergence on an interactionist approach: Emotion expression and impression are determined by an interaction of psychobiological and sociocultural factors. In addition, there may be epochal factors, i.e., diachronic development of expression modes due to *Zeitgeist*, fashion, and other influences changing over time (Hollywood movies, Facebook). The question of whether expression and impression of emotion are either universal or culturally specific in terms of a dichotomy is therefore moot, as science has progressed to a higher level of understanding (Elfenbein & Ambady, 2002; see also Mesquita et al., 1997). However, as the review has shown, this convergence of opinion has not yet led to a radical paradigm shift with respect to the empirical work. In the future, much effort needs to be directed at (a) a clear theoretical underpinning of the research, resulting in clearly operationalized and justified hypotheses; (b) the adoption of a process model that contains both the encoding and decoding aspects, as well as additional subprocesses (as sketched in the Brunswikian lens model); (c) the use of sampling procedures that avoid convenience sampling and attempt a theoretically justified selection of salient cultural settings to be compared.

Future work should also endeavor to uncover and/or validate the mechanism underlying emotion expression and perception. Except for Darwin, there has been surprisingly little concern about the nature of the presumed mechanism underlying expression. With respect to emotion, two major competing positions specify testable predictions: (a) discrete or basic emotion theories (Ekman, 1972; Izard, 1971; Tomkins, 1962), postulating that affect programs for basic emotions, such as anger, fear, sadness, and joy, produce prototypical response configurations that include emotion-specific patterns of expressions; and (b) componential emotion models (Scherer, 1985, 1988, 1992, 2001; Smith & Scott, 1997), postulating that the individual elements of facial expression are determined by appraisal results and their effects on motor behavior. Frijda and Tcherkassof's (1997) proposal to view facial expression as modes of action tendencies is consistent with appraisal theory as motivational factors mediate the production of motor behavior.

Although both models described above adopt a functional approach based on Darwin, they differ in the scope and level of predetermination of the predicted expression patterns. Past research has shown little evidence for frequent occurrence of the well-formed, prototypical, and highly emotion-specific expression patterns or configurations that would be expected as the result of affect programs (see Scherer & Ellgring, 2007a). Experimental research on the notion of componential emotion theorists, assuming that appraisal mediated through motivational action tendencies elicits and differentiates expression, has barely started. It will be an important task for the future, essential for further work on cultural influences, to design studies that allow a critical test between these two competing theories.

Importantly, there are modality-specific factors to be taken into account for facial, vocal, and gestural expressions. Furthermore, the enormous role of situational context (e.g., Righart & de Gelder, 2008), as well as of strategic manipulation and regulation attempts (Scherer & Bänziger, 2010), needs finally to be acknowledged in research design and at least accounted for by mediation analyses. Last, but not least, given the remarkable differences between static and dynamic emotion representations and the central role of expression modality, we strongly recommend the use of multimodal dynamic stimuli in further research. In conclusion, despite the venerable age of research on emotion expression we may not have progressed as much as would be desirable given the importance for the phenomenon for social interaction, media effects, diagnosis of affect disturbances, and a host of other important applied issues. It is to be hoped that in addition to the recent trend to focus on naturally occurring expression in the field, there will also be more theoretically motivated and experimentally controlled studies—particularly across cultures—that will advance further our understanding of the psychobiological and sociocultural factors and mechanisms underlying the expression of emotion.

Manuscript received April 2011
Revised manuscript accepted June 2011

## REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral & Cognitive Neuroscience Reviews, 1*, 21–62.

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior, 33*, 17–34.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face. *Psychological Science, 16*, 403–410.

Anolli, L., Wang, L., Mantovani, F., & De Toni, A. (2008). The voice of emotion in Chinese and Italian young adults. *Journal of Cross-Cultural Psychology, 39*, 565–598.

Argyle, M., & Kendon, A. (1967). The experimental analysis of social performance. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 3, pp. 55–98). New York, NY: Academic Press.

Arnold, M. B. (1960). *Emotion and personality: Vol. 1. Psychological aspects*. New York, NY: Columbia University Press.

Ashton-James, C., Maddux, W., Galinsky, A., & Chartrand, T. (2009). Who I am depends on how I feel: The role of affect in the expression of culture. *Psychological Science, 20*, 340–346.

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6*, 219–224.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in verbal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The multimodal emotion recognition test (MERT). *Emotion, 9*, 691–704.

Bänziger, T., Mortillaro, M., & Scherer, K. R. (in press). Introducing a new 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 K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271–294). Oxford, UK: Oxford University Press.

Barfield, T. (1997). *The dictionary of anthropology*. Oxford, UK: Blackwell.

Beaupré, M., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology, 36*, 355–370.

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

Boucher, J., & Carlson, G. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology, 11*, 263–280.

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion, 15*(1), 57–79.

Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1). Chicago, IL: University of Chicago Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Buck, R. (1984). *The communication of emotion*. New York, NY: Guilford.

Camras, L. A., Bakeman, R., Chen, Y., Norris, K., & Cain, T. R. (2006). Culture, ethnicity, and children's facial expressions: A study of European American, mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion, 6*, 103–114.

Camras, L. A., Meng, Z., Ujiie, T., Dharamsi, S., Miyake, K., Oster, H., et al. (2002). Observing emotion in infants: Facial expression, body behavior and rater judgments of responses to an expectancy-violating event. *Emotion, 2*, 178–193.

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development, 59*, 604–616.

Cohen, I., Sebe, N., Garg, A., Chen, L., & Huang, T. (2003). Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding, 91*, 160–187.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242*, 126–135.

Cowie, R., Douglas-Cowie, E., Sneddon, I., McRorie, M., Hanratty, J., McMahon, E., et al. (2010). Induction techniques developed to illuminate relationships between signs of emotion and their context, physical and social. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 295–307). Oxford, UK: Oxford University Press.

Dael, N., Mortillaro, M., & Scherer, K. R. (in press-a). The body action and posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior*.

Dael, N., Mortillaro, M., & Scherer, K. R. (in press-b). Emotion expression in body action and posture. *Emotion*.

Dailey, M. N., Joyce, C., Lyons, M. J., Kamachi, M., Ishi, H., Gyoba, J., et al. (2010). Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion, 10*, 874–893.

Darwin, C. (1998). *The expression of emotion in man and animals* (3rd ed., P. Ekman, Ed.). Oxford, UK: Oxford University Press. (Original work published 1872).

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication, 40*, 33–60.

Ducci, L., Arcuri, L., Georgis, T., & Sineshaw, T. (1982). Emotion recognition in Ethiopia: The effect of familiarity with Western culture on accuracy of recognition. *Journal of Cross-Cultural Psychology, 13*, 340–351.

Eibl-Eibesfeldt, I. (1989). *Human ethology*. New York, NY: Aldine de Gruyter.

Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology, 2*, 726–735.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska Symposium on Motivation* (pp. 207–282). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In H. Wagner, & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 143–164). Chichester, UK: Wiley.

Ekman, P. (1992a). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P. (1992b). Facial expressions of emotion: New findings, new questions. *Psychological Science, 3*, 34–38.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115*, 268–287.

Ekman, P. (2003). *Emotions revealed*. New York, NY: Times Books.

Ekman, P., & Friesen, W. (1967). Head and body cues in the judgment of emotion: A reformulation. *Perceptual and Motor Skills, 24*, 711–724.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior—Categories, origins, usage, and coding. *Semiotica, 1*, 49–98.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124–129.

Ekman, P., & Friesen, W. (1975). *Unmasking the face*. Cambridge, UK: Malor Books.

Ekman, P., & Friesen, W. (1978). *Manual for the Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W., & Hager, J. C. (2002). *Facial Action Coding System: Investigator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.

Ekman, P., Friesen, W., & Tomkins, S. (1971). Facial affect scoring technique: A first validity study. *Semiotica, 3*, 37–58.

Ekman, P., & Rosenberg, E. (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)* (2nd ed.). New York, NY: Oxford University Press.

Ekman, P., Sorenson, R., & Friesen, W. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*, 86–88.

Elfenbein, H. (2006). Learning in emotion judgments: Training and the cross-cultural understanding of facial expressions. *Journal of Nonverbal Behavior, 30*, 21–36.

Elfenbein, H., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203–235.

Elfenbein, H., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science, 12*, 6.

Elfenbein, H., Lévesque, M., Beaupré, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7*, 131–146.

Elfenbein, H., & Mandal, M. (2004). Hemifacial differences in the in-group advantage in emotion recognition. *Cognition and Emotion, 18*, 613–629.

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

Exline, R. V., & Winters, L. C. (1965). Affective relations and mutual glances in dyads. In S. S. Tomkins, & C. Izard (Eds.), *Affect, cognition, and personality*. New York, NY: Springer.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology, 80*, 75–85.

Fridlund, A. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Frijda, N. H., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell, & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). Cambridge, UK: Cambridge University Press.

Gaspar, A. (2006). Universals and individuality in facial behavior—Past and future of an evolutionary perspective. *Acta Ethologica, 9*, 1–14.

Gifford, R. (1994). A lens framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology, 66*, 398–412.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71*, 878–894.

Goudbeek, M., & Scherer, K. R. (2010). Beyond arousal: Valence and potency/control in the vocal expression of emotion. *Journal of the Acoustical Society of America, 128*, 1322–1336.

Graham, C. R., Hamblin, A., & Feldstein, S. (2001). Recognition of emotion in English voices by speakers of Japanese, Spanish, and English. *International Review of Applied Linguistics in Language Teaching, 39*, 19–37.

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition & Emotion, 13*, 225–266.

Hammond, K., & Stewart, T. (Eds.). (2001). *The essential Brunswik: Beginnings, explications, applications*. New York, NY: Oxford University Press.

Hawk, S., Van Kleef, G., Fischer, A., & Van der Schalk, J. (2009). Worth a thousand words: Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion, 9*, 293–305.

Hjortsjö, C. H. (1970). *Man's face and mimic language*. Malmo, Sweden: Nordens Boktryckeri.

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

Izard, C. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*, 288–299.

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

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology, 6*, 1797–1813.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814.

Juslin, P. N., & Scherer, K. R. (2005). Vocal expression of affect. In J. Harrigan, R. Rosenthal, & K. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 65–135). Oxford, UK: Oxford University Press.

Kappas, A., Hess, U., & Scherer, K. R. (1991). Voice and emotion. In R. S. Feldman, & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 200–238). Cambridge, UK: Cambridge University Press.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22–63.

Kleiner, F. S., & Mamiya, C. J. (2006). *Gardner's art through the ages: The Western perspective: Vol. 1* (12th ed., pp. 11–12). Belmont, CA: Wadsworth.

Kleinsmith, A., De Silva, P., & Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers, 18*, 1371–1389.

Kraus, M., Côté, S., & Keltner, D. (2010). Social class, contextualism, and empathic accuracy. *Psychological Science, 21*, 1716–1723.

Krauss, R., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology, 40*, 312–320.

Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion, 9*, 807–820.

Krumhuber, E., Tamarit, L., Roesch, E. B., & Scherer, K. R. (in press). FACSGen 2.0 animation software: Generating realistic 3D FACS-validated stimuli for facial expression research. *Emotion*.

Kudoh, T., & Matsumoto, D. (1985). Cross-cultural examination of the semantic dimensions of body postures. *Journal of Personality and Social Psychology, 48*, 1440–1446.

Ladd, D., Silverman, K., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. *Journal of the Acoustical Society of America, 78*, 435–444.

Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011). Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer Speech and Language, 25*, 84–104.

Lazarus, R. S. (1968). Emotions and adaptation: Conceptual and empirical relations. In W. J. Arnold (Ed.), *Nebraska Symposium on Motivation* (Vol. 16, pp. 175–270). Lincoln, NE: University of Nebraska Press.

Lazarus, R. S., Coyne, J. C., & Folkman, S. (1984). Cognition, emotion and motivation: The doctoring of 'humpty-dumpty'. In K. R. Scherer, & P. Ekman (Eds.), *Approaches to emotion* (pp. 221–237). Hillsdale, NJ: Lawrence Erlbaum Associates.

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development, 13*, 343–354.

Leu, J., Mesquita, B., Ellsworth, P., ZhiYong, Z., Huijuan, Y., Buchtel, E., et al. (2010). Situational differences in dialectical emotions: Boundary conditions in a cultural comparison or North Americans and East Asians. *Cognition & Emotion, 24*, 419–435.

Lutz, C., & White, G. (1986). The anthropology of emotions. *Annual Review of Anthropology, 15*, 405–436.

Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child Development, 53*, 991–1003.

Mandal, M., & Ambady, N. (2004). Laterality of facial expressions of emotion: Universal and culture-specific influences. *Behavioural Neurology, 15*, 23–34.

Markham, R., & Wang, L. (1996). Recognition of emotion by Chinese and Australian children. *Journal of Cross-Cultural Psychology, 27*, 613–643.

Marsh, A., Elfenbein, H., & Ambady, N. (2003). Nonverbal "accents": Cultural differences in facial expressions of emotion. *Psychological Science, 14*, 373–376.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review, 3*, 215–221.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, K., & van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology, 94*, 365–381.

Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology, 20*, 92–105.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion, 14*, 195–214.

Matsumoto, D. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion, 17*, 107–123.

Matsumoto, D., & Assar, M. (1992). The effects of language on judgments of universal facial exressions of emotion. *Journal of Nonverbal Behavoir, 16*, 85–99.
Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., Paul, S., et al. (2002). American–Japanese cultural differences in judgements of emotional expressions of different intensities. *Cognition and Emotion, 16*, 721–747.
Matsumoto, D., & Ekman, P. (1989). American–Japanese cultural differences in intensity ratings of facial expressions of emotion I. *Motivation and Emotion, 13*, 143–157.
Matsumoto, D., Kasri, F., & Kooken, K. (1999). American–Japanese cultural differences in judgements of expression intensity and subjective experience. *Cognition and Emotion, 13*, 201–218.
Matsumoto, D., & Kudoh, T. (1993). American-Japanese cultural differences in attributions of personality based on smiles. *Journal of Nonverbal Behavior, 17*, 231–243.
Matsumoto, D., Olide, A., Schug, J., Willingham, B., & Callan, M. (2009). Cross-cultural judgments of spontaneous facial expressions of emotion. *Journal of Nonverbal Behavior, 33*, 213–238.
McAndrew, F. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology, 17*, 211–224.
Mead, M. (1928). *Coming of age in Samoa*. New York, NY: William Morrow.
Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16518–16523.
Mehrabian, A. (1968a). The effect of context on judgments of speaker attitude. *Journal of Personality, 36*, 21–32.
Mehrabian, A. (1968b). Inference of attitudes from the posture, orientation, and distance of a communicator. *Journal of Consulting and Clinical Psychology, 32*, 296–308.
Mehrabian, A. (1972). *Nonverbal communication*. Chicago, IL: Aldine-Atherton.
Mehrabian, A. (2007). Communication without words. In D. Mortensen (Ed.), *Communication theory* (pp. 47–57). New Brunswick, NJ: Transaction.
Mehrabian, A., & Friar, J. (1969). Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology, 33*, 330–336.
Mehu, M., Mortillaro, M., & Scherer, K. R. (2011). *Reliable facial muscles activation enhances the recognizability and credibility of emotional expression*. Manuscript submitted for publication.
Mesquita, B., Frijda, N., & Scherer, K. R. (1997). Culture and emotion. In J. W. Berry, & P. R. Dasen (Eds.), *Handbook of cross-cultural psychology: Vol. 2. Basic processes and human development* (pp. 255–297). Needham Heights, MA: Allyn & Bacon.
Morris, D. (1977). *Manwatching: A field guide to human behavior*. New York, NY: H.N. Abrams.
Morris, M., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*, 949–971.
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.
Murray, I. R., & Arnott, J. L. (1993). Toward a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*, 1097–1108.
Naab, P., & Russell, J. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion, 7*, 736–744.
Niit, T., & Valsiner, J. (1977). Recognition of facial expressions: An experimental investigation of Ekman's model. *Acta et Commentationes Universitatis Tarwensis, 429*, 85–107.
Osgood, C. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology, 7*, 1–30.
Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science, 283*, 1272–1273.
Patel, S., Scherer, K. R., Sundberg, J., & Björkner, E. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology, 87*, 93–98.
Pell, M., Paulmann, S., Dara, C., Allasseri, A., & Kotz, S. (2009). Factors in the recognition of vocal expression emotions: A comparison of four languages. *Journal of Phonetics, 37*, 417–435.
Reeve, J., & Nix, G. (1997). Expressing intrinsic motivation through acts of exploration and facial displays of interest. *Motivation and Emotion, 21*, 237–250.
Ricci-Bitti, P. E., Brighetti, G., Garotti, P. L., & Boggi-Cavallo, P. (1989). Is contempt expressed by pancultural facial movements? In J. P. Forgas, & J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 329–339). Amsterdam, The Netherlands: Elsevier.
Righart, R., & de Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 264–272.
Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman, & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 239–281). New York, NY: Cambridge University Press.
Rinn, W. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95*, 52–77.
Riskind, J. H. (1984). They stoop to conquer: Guiding and self-regulatory functions of physical posture after success and failure. *Journal of Personality and Social Psychology, 47*, 479–493.
Roach, J. (1985). *The player's passion: Studies in the science of acting*. Newark, NJ: University of Delaware Press.
Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401.
Roesch, E. B., Tamarit, L., Reveret, L., Grandjean, D., Sander, D., & Scherer, K. R. (2010). FACSGen: A tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *Journal of Nonverbal Behavior, 35*, 1–16.
Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial

expressions of emotion. *Motivation & Emotion, 19*(2), 111–138.

Ross, E. D., Edmondson, J. A., & Seibert, G. B. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics, 14*, 283–302.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115*, 102–141.

Russell, J., & Fernandez-Dols, J. (Eds.). (1997). *The psychology of facial expression*. New York, NY: Cambridge University Press.

Russell, J., Suzuki, N., & Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expression. *Motivation and Emotion, 17*, 337–351.

Saha, G. B. (1973). Judgment of facial expression of emotion in nunhuman primates. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 11–89). New York, NY: Academic Press.

Sauter, D., Eisner, F., Ekman, P., & Scott, S. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 2408–2412.

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology, 8*, 467–487.

Scherer, K. R. (1979). Non-linguistic vocal indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions in personality and psychopathology* (pp. 495–529). New York, NY: Plenum Press.

Scherer, K. R. (1985). Vocal affect signaling: A comparative approach. In J. S. Rosenblatt et al. (Eds.), *Advances in the study of behavior* (Vol. 15, pp. 189–244). New York, NY: Academic Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143–165.

Scherer, K. R. (1988). On the symbolic functions of vocal affect expression. *Journal of Language and Social Psychology, 7*, 79–100.

Scherer, K. R. (1992). What does facial expression express? In K. Strongman (Ed.), *International review of studies on emotion* (Vol. 2, pp. 139–165). Chichester, UK: Wiley.

Scherer, K. R. (1994). Affect bursts. In S. H. M. van Goozen, & N. E. Van de Poll (Eds.), *Emotions: Essays on emotion theory* (pp. 161–193). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology, 73*, 902–922.

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

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227–256.

Scherer, K. R. (2011). *Vocal markers of emotion: Comparing induction and acting elicitation*. Manuscript submitted for publication.

Scherer, K. R., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92.

Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 166–178). Oxford, UK: Oxford University Press.

Scherer, K. R., & Brosch, T. (2009). Culture-specific appraisal biases contribute to emotion dispositions. *European Journal of Personality, 23*, 265–288.

Scherer, K. R., & Ceschi, G. (2000). Studying affective communication in the airport: The case of lost baggage claims. *Personality and Social Psychology Bulletin, 26*, 327–339.

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

Scherer, K. R., & Ellgring, H. (2007b). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion, 7*(1), 158–171.

Scherer, K. R., & Grandjean, D. (2008). Inferences from facial expressions of emotion have many facets. *Cognition and Emotion, 22*(5), 789–801.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. Goldsmith (Eds.), *Handbook of the affective sciences* (pp. 433–456). New York, NY: Oxford University Press.

Scherer, K. R., Johnstone, T., Klasmeyer, G., & Bänziger, T. (2000). Can automatic speaker verification be improved by training the algorithms on emotional speech? *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP, 2000)*, Beijing, China.

Scherer, K. R., & Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*, 331–346.

Scherer, K. R., & Wallbott, H. G. (1985). Analysis of nonverbal behavior. In T. A. van Dijk (Ed.), *Handbook of discourse analysis* (pp. 199–230). London, UK: Academic Press.

Scherer, K. R., Wallbott, H. G., & Scherer, U. (1979). Methoden zur Klassifikation von Bewegungsverhalten: Ein funktionaler Ansatz [Methods for the classification of movement behavior: A functional approach]. *Zeitschrift für Semiotik, 1*, 177–192.

Scherer, K. R., Zentner, M., & Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion, 4*, 389–402.

Schimmack, U. (1996). Cultural influences on the recognition of emotion by facial expressions: Individualistic or Caucasian cultures? *Journal of Cross-Cultural Psychology, 27*, 37–50.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61*, 8.

Shackman, J. E., & Pollak, S. D. (2005). Experiential influences on multimodal perception of emotion. *Child Development, 76*, 1116–1126.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for general causal inference*. Boston, MA: Houghton Mifflin.

Simon, D., Craig, K. D., Gosselin, F., Belin, P., & Rainville, P. (2008). Recognition and discrimination of prototypical dynamic expressions of pain and emotions. *Pain, 135*, 55–64.

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9*, 838–846.

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

Sundberg, J., Patel, S., Björkner, E., & Scherer, K. R. (2011). Interdependencies among voice source parameters in emotional speech. *IEEE Transactions on Affective Computing, 2*, 162–174.

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., & Adam, J. (2007). Facial expressions of emotions: A methodological contribution to the study of spontaneous and dynamic emotional faces. *European Journal of Social Psychology, 37*, 1325–1345.

Tcherkassof, A., & Suremain, F. (2005). Burkina Faso and France: A cross-cultural study of the judgment of action readiness in facial expressions of emotion. *Psychologia, 48*, 317–334.

Thompson, W., & Balkwill, L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*, 407–424.

Tomkins, S. (1962). *Affect imagery consciousness: Volume 1, The positive affects*. London, UK: Tavistock.

Van Bezooijen, R., Otto, S., & Heenan, T. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology, 14*, 387–406.

Wagner, H. (1990). The spontaneous facial expression of differential positive and negative emotions. *Motivation and Emotion, 14*, 27–43.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17*, 3–28.

Wagner, H., Lewis, H., Ramsay, S., & Krediet, I. (1992). Prediction of facial displays from knowledge of norms of emotional expressiveness. *Motivation and Emotion, 16*, 347–362.

Wagner, H., MacDonald, C., & Manstead, A. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology, 50*, 737–743.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*, 437–456.

Wallbott, H. G. (1985). Hand movement quality: A neglected aspect of nonverbal behavior in clinical judgment and person perception. *Journal of Clinical Psychology, 41*(3), 345–359.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*(6), 879–896.

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51*, 690–699.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology, 78*, 105–119.

Williams, F., & Sundene, B. (1965). Dimensions of recognition: Visual vs. vocal expression of emotion. *Educational Technology Research and Development, 13*, 44–52.

Winkelmayer, R., Exline, R., Gottheil, E., & Paredes, A. (1978). The relative accuracy of U.S., British, and Mexican raters in judging the emotional displays of schizophrenic and normal U.S. women. *Comparative and General Pharmacology, 34*, 600–608.

Yik, M., Meng, Z., & Russell, J. (1998). Adults' freely produced emotion labels for babies' spontaneous facial expressions. *Cognition and Emotion, 12*, 723–730.

Yik, M., & Russell, J. (1999). Interpretation of faces: A cross-cultural study of a prediction from Fridlund's theory. *Cognition and Emotion, 13*, 93–104.

Young, S., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology, 99*, 964–977.

## APPENDIX

**TABLE A1**
Studies examining the expression and perception of emotion in static facial expressions (photos)

| *Citation* | *Encoder culture* | *Decoder culture* | *Answer alternatives* | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* |
|---|---|---|---|---|---|---|---|---|---|---|
| Bänziger et al. (2009) | Swiss | Swiss | 10 | 91 | – | 57 | 80 | 85 | 52 | 73 |
| Beaupré & Hess (2005) | African | African | 9 | 78 | – | 24 | 35 | 42 | 57 | 47 |
| Beaupré & Hess (2005) | Chinese | African | 9 | 73 | – | 31 | 20 | 54 | 43 | 44 |
| Beaupré & Hess (2005) | African | Chinese | 9 | 76 | – | 40 | 30 | 47 | 41 | 47 |
| Beaupré & Hess (2005) | Chinese | Chinese | 9 | 70 | – | 42 | 27 | 54 | 46 | 48 |
| Beaupré & Hess (2005) | African | Canadian | 9 | 68 | – | 52 | 36 | 55 | 51 | 52 |
| Beaupré & Hess (2005) | Chinese | Canadian | 9 | 70 | – | 65 | 27 | 55 | 49 | 53 |
| Beaupré & Hess (2005) | Canadian | African | 9 | 78 | – | 32 | 26 | 43 | 47 | 45 |
| Beaupré & Hess (2005) | Canadian | Chinese | 9 | 67 | – | 31 | 27 | 47 | 47 | 44 |
| Beaupré & Hess (2005) | Canadian | Canadian | 9 | 63 | – | 52 | 37 | 59 | 51 | 52 |
| Biehl et al. (1997) | American & Japanese | Japanese | 7 | 98 | 92 | 72 | 55 | 75 | 64 | 76 |
| Biehl et al. (1997) | American & Japanese | Sumatran | 7 | 99 | 89 | 81 | 57 | 76 | 79 | 80 |
| Biehl et al. (1997) | American & Japanese | Vietnamese | 7 | 99 | 92 | 80 | 67 | 58 | 81 | 80 |
| Biehl et al. (1997)[d] | American & Japanese | American | 7 | 98 | 92 | 92 | 79 | 81 | 84 | 88 |
| Biehl et al. (1997) | American & Japanese | Hungary | 7 | 98 | 94 | 83 | 74 | 84 | 87 | 87 |
| Biehl et al. (1997)[a] | American & Japanese | Polish | 7 | 98 | 89 | 88 | 69 | 83 | 85 | 85 |
| Boucher & Carlson (1980) | Malaysian | Malaysian | 6 | 96 | 73 | 65 | 25 | 52 | 50 | 60 |
| Boucher & Carlson (1980) | Malaysian | American | 6 | 98 | 86 | 67 | 49 | 57 | 59 | 69 |
| Boucher & Carlson (1980) | American | Malaysian | 6 | 95 | 67 | 68 | 66 | 67 | 49 | 69 |
| Boucher & Carlson (1980) | American | Malaysian | 6 | 90 | 80 | 74 | 53 | 85 | 73 | 76 |
| Boucher & Carlson (1980) | American | American | 6 | 96 | 86 | 79 | 91 | 87 | 71 | 85 |
| Ducci et al. (1982) | American | Ethiopian | 7 | 87 | 51 | 52 | 59 | 55 | 37 | 57 |
| Ekman & Friesen (1971) | American | New Guinean | 7 | 92 | 98 | 81 | 88 | 85 | 90 | 89 |
| Ekman (1972) | American | Japanese | 6 | 87 | 87 | 80 | 71 | 82 | 63 | 78 |
| Ekman (1972) | American | New Guinean | 2 | 98 | 89 | 77 | 72 | 91 | 70 | 83 |
| Ekman (1972) | American | New Guinean | 2 | 92 | 98 | 81 | 93 | 85 | 90 | 90 |
| Ekman (1972) | American | Argentine | 6 | 94 | 93 | 88 | 68 | 79 | 72 | 82 |
| Ekman (1972) | American | Brazilian | 6 | 92 | 81 | 87 | 77 | 86 | 68 | 82 |
| Ekman (1972) | American | Chilean | 6 | 90 | 88 | 91 | 78 | 85 | 76 | 85 |
| Ekman (1972) | American | American | 6 | 97 | 91 | 87 | 88 | 83 | 68 | 86 |
| Ekman et al. (1969) | American | Japanese | 6 | 87 | 87 | 74 | 71 | 82 | 63 | 77 |
| Ekman et al. (1969) | American | Malaysian | 6 | 92 | 36 | 52 | 40 | | 64 | 57 |
| Ekman et al. (1969) | American | New Guinean | 6 | 90 | – | 55 | 50 | 36 | 53 | 57 |
| Ekman et al. (1969) | American | American | 6 | 97 | 91 | 73 | 88 | 82 | 69 | 83 |
| Ekman et al. (1969) | American | Brazilian | 6 | 97 | 82 | 82 | 77 | 86 | 82 | 84 |
| Ekman et al. (1987) | American | Hong Kong | 7 | 92 | 91 | 91 | 84 | 65 | 73 | 83 |
| Ekman et al. (1987) | American | Japanese | 7 | 90 | 94 | 87 | 65 | 60 | 67 | 77 |
| Ekman et al. (1987) | American | Sumatran | 7 | 69 | 78 | 91 | 70 | 70 | 70 | 75 |
| Ekman et al. (1987) | American | Turkish | 7 | 87 | 90 | 76 | 76 | 74 | 79 | 80 |
| Ekman et al. (1987) | American | American | 7 | 95 | 92 | 92 | 84 | 86 | 81 | 88 |
| Ekman et al. (1987) | American | Estonian | 7 | 90 | 94 | 86 | 91 | 71 | 67 | 83 |
| Ekman et al. (1987) | American | German | 7 | 93 | 87 | 83 | 86 | 61 | 71 | 80 |
| Ekman et al. (1987) | American | Greek | 7 | 93 | 91 | 80 | 74 | 77 | 77 | 82 |
| Ekman et al. (1987) | American | Italian | 7 | 97 | 92 | 81 | 82 | 89 | 72 | 86 |
| Ekman et al. (1987) | American | Scottish | 7 | 98 | 88 | 86 | 86 | 79 | 84 | 87 |
| Elfenbein (2006) | Chinese | Chinese | 5 | – | 85 | 74 | 74 | – | 67 | 75 |
| Elfenbein (2006) | Chinese | American | 5 | – | 85 | 71 | 81 | – | 71 | 77 |
| Elfenbein (2006) | American | Chinese | 5 | – | 90 | 81 | 61 | – | 75 | 77 |
| Elfenbein (2006) | American | American | 5 | – | 90 | 80 | 64 | – | 89 | 81 |
| Elfenbein et al. (2007) | African | Gabonese | 10 | 86 | – | 45 | 50 | 35 | 60 | 55 |
| Elfenbein et al. (2007) | Gabonese | Gabonese | 10 | 68 | 50 | 38 | 18 | 35 | 33 | 40 |
| Elfenbein et al. (2007) | Gabonese | Canadian | 10 | 78 | 54 | 38 | 19 | 40 | 26 | 43 |
| Elfenbein et al. (2007) | Canadian | Gabonese | 10 | 70 | 45 | 29 | 28 | 30 | 39 | 40 |
| Elfenbein et al. (2007) | Caucasian | Gabonese | 10 | 74 | – | 36 | 49 | 29 | 66 | 51 |
| Elfenbein et al. (2007) | Canadian | Canadian | 10 | 80 | 56 | 39 | 34 | 54 | 47 | 52 |
| Elfenbein et al. (2007) | Canadian | Canadian | 10 | 86 | – | 66 | 63 | 52 | 78 | 69 |
| Elfenbein et al. (2007) | African | Canadian | 10 | 90 | – | 76 | 73 | 71 | 65 | 75 |

(*continued*)

**TABLE A1.** Continued.

| *Citation* | *Encoder culture* | *Decoder culture* | *Answer alternatives* | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* |
|---|---|---|---|---|---|---|---|---|---|---|
| Izard (1971) | American & French | African | 8 | 68 | 49 | 32 | 49 | 55 | 51 | 51 |
| Izard (1971) | American | Japanese | 8 | 94 | 79 | 67 | 58 | 56 | 57 | 69 |
| Izard (1971) | American & French | American | 8 | 97 | 91 | 74 | 76 | 83 | 89 | 85 |
| Izard (1971) | American & French | American | 8 | 90 | 89 | 60 | 60 | 49 | 66 | 69 |
| Izard (1971) | American & French | British | 8 | 96 | 81 | 75 | 67 | 85 | 82 | 81 |
| Izard (1971) | American & French | British | 8 | 80 | 80 | 64 | 59 | 46 | 57 | 64 |
| Izard (1971) | American & French | French | 8 | 95 | 84 | 71 | 84 | 79 | 92 | 84 |
| Izard (1971) | American & French | French | 8 | 85 | 65 | 59 | 62 | 47 | 54 | 62 |
| Izard (1971) | American & French | German | 8 | 98 | 86 | 67 | 84 | 73 | 83 | 82 |
| Izard (1971) | American & French | Greek | 8 | 94 | 80 | 55 | 68 | 88 | 80 | 78 |
| Izard (1971) | American & French | Greek | 8 | 80 | 56 | 50 | 71 | 49 | 47 | 59 |
| Izard (1971) | American & French | Swedish | 8 | 97 | 81 | 72 | 89 | 88 | 82 | 85 |
| Izard (1971) | American & French | Swiss | 8 | 97 | 86 | 70 | 68 | 78 | 92 | 82 |
| Markham & Wang (1996)[b] | American & Chinese | Chinese | 2 | 100 | 92 | 90 | 84 | 80 | 88 | 89 |
| Markham & Wang (1996)[b] | American & Chinese | Australian | 2 | 95 | 73 | 70 | 75 | 55 | 65 | 72 |
| Matsumoto & Assar (1992)[c] | Japanese | Indian | 5 | 99 | 90 | 94 | 81 | – | 91 | 91 |
| Matsumoto & Assar (1992)[d] | Japanese | Indian | 5 | 99 | 90 | 90 | 74 | – | 89 | 88 |
| Matsumoto et al. (2002) | American & Japanese | Japanese | 9 | 89 | 89 | 48 | – | – | 55 | 70 |
| Matsumoto et al. (2002) | American & Japanese | American | 9 | 93 | 83 | 66 | – | – | 68 | 78 |
| McAndrew (1986) | American | Malaysian | 6 | 100 | 95 | 100 | 67 | 98 | 86 | 91 |
| McAndrew (1986) | American | American | 6 | 100 | 93 | 88 | 68 | 93 | 90 | 89 |
| Naab & Russell (2007) | New Guinean | American | 12 | 33 | 34 | 46 | 17 | 25 | 28 | 30 |
| Niit & Valsiner (1977) | American | Kirghizian | 7 | 89 | 71 | 89 | 51 | 86 | 47 | 72 |
| Niit & Valsiner (1977) | American | Estonian | 7 | 88 | 83 | 85 | 60 | 89 | 78 | 81 |
| Russell et al. (1993) | American | Japanese | 7 | 84 | 94 | 80 | 14 | 56 | 48 | 63 |
| Russell et al. (1993) | American | Greek | 7 | 92 | 55 | 75 | 87 | 68 | 63 | 73 |
| Russell et al. (1993) | American & Japanese | Canadian | 7 | 100 | 96 | 70 | 62 | 66 | 78 | 79 |
| Yik & Russell (1999) | American | Chinese | 10 | 90 | 63 | 78 | 33 | 53 | 48 | 61 |
| Yik & Russell (1999) | American | Japanese | 10 | 92 | 75 | 73 | 53 | 37 | 57 | 65 |
| Yik & Russell (1999) | American | Canadian | 10 | 88 | 72 | 78 | 67 | 60 | 67 | 72 |
| Yik et al. (1998) | Chinese | Chinese | n/a | 74 | 7 | 26 | 20 | 2 | 5 | 22 |
| Yik et al. (1998) | Chinese | Japanese | 6 | 72 | 33 | 63 | 26 | 4 | 14 | 35 |
| Yik et al. (1998) | Chinese | Canadian | 6 | 77 | 32 | 48 | 28 | 4 | 11 | 33 |

A dash indicates that no data are available. N/a = not applicable. All data are expressed as percentages. Individual studies listed alphabetically by author name. [a]This study was not used in the compilation for Table 2A as encoder cultures were not separated. [b]Markham & Wang (1996) examined emotion perception in three groups of children, aged 4, 6, and 8 years; the results reported here are averaged over those three groups; the data are reported in graphical form only in the original publication. [c]Experiment conducted in English (participants were English–Hindi bilingual). [d]Experiment conducted in Hindi (participants were English–Hindi bilingual).

**TABLE A2**

Studies examining the expression and perception of emotion in dynamic facial expression (video)

| *Citation* | *Encoder culture* | *Decoder culture* | *Answer alternatives* | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* |
|---|---|---|---|---|---|---|---|---|---|---|
| Bänziger & Scherer (2010) | Swiss | Swiss | 15 | 75 | 27 | 58 | 67 | 43 | 79 | 58 |
| Bänziger & Scherer (2010) | Swiss | Swiss | 15 | 76 | 53 | 68 | 75 | 71 | 76 | 70 |
| Bänziger et al. (2009) | Swiss | Swiss | 10 | 93 | – | 72 | 68 | 72 | 83 | 78 |
| Ekman (1972)[b] | New Guinean | American | 6 | 73 | 27 | 68 | 18 | 46 | 51 | 47 |
| Hawk et al. (2009) | Dutch | Dutch | 10 | 90 | 76 | 86 | 70 | 83 | 80 | 81 |
| Wallbott & Scherer (1986) | German | German | 4 | 62 | 45 | 63 | – | – | 84 | 64 |
| Winkelmayer et al. (1978)[a] | American | American | 3 | 43 | – | 44 | – | – | 36 | 41 |
| Winkelmayer et al. (1978)[a] | American | British | 3 | 46 | – | 42 | – | – | 36 | 41 |
| Winkelmayer et al. (1978)[a] | American | Mexican | 3 | 42 | – | 35 | – | – | 32 | 36 |

A dash indicates that no data are available. All data are expressed as percentages. Individual studies listed alphabetically by author name.

**TABLE A3**
Studies examining the expression and perception of emotion in vocal information (audio)

| *Citation* | *Encoder culture* | *Decoder culture* | *Language content* | *Answer alternatives* | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Banse & Scherer (1996) | German | German | N | 14 | 45 | – | 64 | 59 | 15 | 70 | 51 |
| Bänziger & Scherer (2010) | Swiss | Swiss | N | 15 | 49 | 26 | 50 | 57 | 10 | 71 | 44 |
| Bänziger & Scherer (2010) | Swiss | Swiss | N | 15 | 60 | 40 | 40 | 59 | 55 | 62 | 53 |
| Bänziger et al. (2009) | Swiss | Swiss | N | 10 | 48 | – | 72 | 66 | 14 | 83 | 57 |
| Graham et al. (2001) | American | Japanese | Y | 8 | 47 | – | 36 | 37 | – | 56 | 44 |
| Graham et al. (2001) | American | American | Y | 7 | 70 | – | 39 | 55 | – | 68 | 58 |
| Graham et al. (2001) | American | Spanish | Y | 7 | 31 | – | 46 | 28 | – | 70 | 44 |
| Hawk et al. (2009) | Dutch | Dutch | N | 10 | 86 | 83 | 97 | 93 | 94 | 80 | 89 |
| Hawk et al. (2009) | Dutch | Dutch | Y | 10 | 49 | 98 | 82 | 52 | 37 | 90 | 68 |
| Scherer et al. (1991) | German | German | N | 5 | 54 | – | 62 | 35 | 18 | 56 | 45 |
| Scherer et al. (1991) | German | German | N | 5 | 81 | – | 91 | 84 | 43 | 72 | 74 |
| Scherer et al. (1991) | German | German | N | 5 | 51 | – | 67 | 41 | 27 | 70 | 51 |
| Scherer et al. (1991) | German | German | N | 5 | 48 | – | 69 | 49 | 21 | 75 | 52 |
| Scherer et al. (2001) | German | Indonesian | N | 5 | 28 | – | 58 | 38 | – | 64 | 47 |
| Scherer et al. (2001) | German | American | N | 5 | 46 | – | 73 | 72 | – | 80 | 68 |
| Scherer et al. (2001) | German | British | N | 5 | 40 | – | 82 | 70 | – | 83 | 69 |
| Scherer et al. (2001) | German | Dutch | N | 5 | 45 | – | 69 | 65 | – | 86 | 66 |
| Scherer et al. (2001) | German | French | N | 5 | 51 | – | 67 | 71 | – | 69 | 65 |
| Scherer et al. (2001) | German | German | N | 5 | 48 | – | 80 | 74 | – | 79 | 70 |
| Scherer et al. (2001) | German | Italian | N | 5 | 39 | – | 68 | 77 | – | 72 | 64 |
| Scherer et al. (2001) | German | Spanish | N | 5 | 30 | – | 71 | 65 | – | 73 | 60 |
| Scherer et al. (2001) | German | Swiss | N | 5 | 55 | – | 71 | 70 | – | 79 | 69 |
| Thompson & Balkwill (2006) | Chinese | Canadian | Y | 4 | 48 | – | 61 | 68 | – | 61 | 60 |
| Thompson & Balkwill (2006) | Japanese | Canadian | Y | 4 | 58 | – | 79 | 33 | – | 48 | 55 |
| Thompson & Balkwill (2006) | Philippine | Canadian | Y | 4 | 49 | – | 98 | 54 | – | 89 | 73 |
| Thompson & Balkwill (2006) | Canadian | Canadian | Y | 4 | 99 | – | 90 | 90 | – | 99 | 95 |
| Thompson & Balkwill (2006) | German | Canadian | Y | 4 | 58 | – | 84 | 53 | – | 76 | 68 |
| van Bezooijen et al. (1983) | Dutch | Japanese | Y | 10 | 20 | 29 | 70 | 19 | 26 | 40 | 34 |
| van Bezooijen et al. (1983) | Dutch | Taiwanese | Y | 10 | 24 | 53 | 53 | 36 | 26 | 47 | 40 |
| van Bezooijen et al. (1983) | Dutch | Dutch | Y | 9 | 76 | 68 | 73 | 51 | 55 | 70 | 66 |
| Wallbott & Scherer (1986) | German | German | N | 4 | 37 | 32 | 57 | | | 65 | 48 |

A dash indicates that no data are available. Y = yes; N = no. All data are expressed as percentages. Individual studies listed alphabetically by author name.

**TABLE A4**
Studies examining the expression and perception of emotion in combined facial and vocal information (audio–video)

| *Citation* | *Encoder culture* | *Decoder culture* | *Language content* | *Answer alternatives* | *Happiness* | *Surprise* | *Sadness* | *Fear* | *Disgust* | *Anger* | *Mean* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bänziger & Scherer (2010) | Swiss | Swiss | N | 15 | 76 | 46 | 67 | 76 | 70 | 83 | 70 |
| Bänziger & Scherer (2010) | Swiss | Swiss | N | 15 | 74 | 60 | 56 | 84 | 98 | 82 | 76 |
| Bänziger et al. (2009) | Swiss | Swiss | N | 10 | 92 | – | 80 | 78 | 72 | 86 | 82 |
| Wallbott & Scherer (1986) | German | German | N | 4 | 50 | 49 | 66 | – | – | 82 | 62 |

A dash indicates that no data are available. N = no. All data are expressed as percentages. Individual studies listed alphabetically by author name.