29

OBSERVATIONAL APPROACHES TO THE MEASUREMENT OF EMOTIONS

Rainer Reisenzein[a], Martin Junge[a], Markus Studtmann[b], and Oswald Huber[c], University of Greifswald[a], Max Planck Institute for Human Development[b], and University of Fribourg[c]

The hallmark of psychology as an empirical science is the reliance on empirical data to support its claims. As traditionally conceived in philosophy and psychology (see, e.g., Brentano, 1874; Wundt, 1896), *empirical data* comprise all kinds of information obtained through experience, both those acquired through the "outer senses" (the sense-organs; i.e., the eyes, ears, etc.) and those acquired through the "inner sense", the self-observation of conscious mental states (also called *introspection*).[1] Common sense suggests that for obtaining empirical information about emotions, both methods—introspective self-observation and external observation—are useful. For example, to acquire empirical information about anger, one can either observe one's own feelings and thoughts during an episode of anger or ask others to report their experiences; or one can watch what other people do, say, and express nonverbally in their face, voice, and body when they are angry. In agreement with common sense, both introspection and external observation have been extensively used in emotion research since the beginnings of psychology as an independent discipline in the late 19th century (e.g., Wundt, 1896). Over the years, both methods have evolved and now exist in several more or less standardized variants. Introspection-based observation methods are today usually referred to as *self-report methods* (see Pekrun & Bühner, 2014). Their best-known incarnation in current emotion psychology is the emotion rating scale, but several other self-report based methods useful for emotion research do exist (e.g., Junge & Reisenzein, 2013). In this chapter, however, interest is on methods of emotion assessment based on external observation.

From the systematic perspective, methods of external observation in psychology comprise all methods that are based on the observation of intersubjectively accessible aspects of a psychological phenomenon (e.g., emotion), with or without the help of special observation instruments (e.g., video cameras, physiological sensors), and with or without making inferences to underlying mental states. Hence, from the systematic perspective, methods of external observation include not only assessments of behavior but also psychophysiological and neurophysiological measurements (see Immordino-Yang & Christodoulou, 2014; Kreibig & Gendolla, 2014). In this chapter, however, we focus on a subset of the methods of external observation commonly called *behavior observation methods* or simply *observational methods.* They can be defined as those methods of external observation that use human observers—or recently, machine substitutes of human observers (see the last part of this chapter)—as measurement devices and as a consequence are restricted to behaviors and events that are observable by humans, although the observers may be (and in fact typically are) asked to draw inferences to underlying mental states.[2] The following description by Wright (1960) captures the essence of behavior observation well: "One gets within seeing or hearing distance [of the target person], observes something about his behavior or situation or both, and then scores, classifies, summarizes, freely interprets, or otherwise does something with the recorded observations" (p. 71). As Wright's description also suggests, behavior observation methods rely primarily on visual and acoustic information, reflecting the fact that sight and hearing are the two most information-rich sensory channels of humans.

FOUNDATIONS OF OBSERVATIONAL APPROACHES TO THE MEASUREMENT OF EMOTIONS

We first describe the commonalities and differences of everyday and systematic behavior observations of emotions and discuss in what sense behavior observation can be regarded as a form of measurement. Then we discuss the kinds and diagnostic value of the cues to emotion that are in principle available to observers. In the central part of the chapter, we describe examples of the three main approaches to the observational measurement of emotions: objective behavior coding systems, theory-based coding methods, and the intuitive observer judgment method. Finally, we make recommendations on the practical implementation of observational emotion measurement and provide some information about recent developments in the field of behavior-based, automatic affect detection.

Everyday Versus Systematic Observations of Emotions

Observations of other people's behaviors (e.g., John smiles) and inferences from these behaviors and the context in which they occur to underlying emotions (e.g., John is amused) are commonly made by all of us in daily life (e.g., Heider, 1958; Malle, 2004; Schneider, Hastorf, & Ellsworth, 1979). Sometimes, we may even have the impression that we do not *make an inference* at all but literally *see* that another person is amused, sad, and so forth, but this shows only that the inference process can become fast and automatic to the point where it appears to be an integral part of perception. The inference of emotions is a special case of the more general phenomenon of *mindreading* or *mental state detection*—the inference of other people's mental states (which include, in addition to emotions, beliefs, desires, intentions, perceptions, and more).[3] Like its special form, emotion inference, mental state detection is routinely performed in everyday life, and the ability to engage in it is an essential component of humans' folk-psychological capacity. Mental state detection is best conceptualized as a process of *multiple-cue based, folk-theory-guided inference to the best explanation* (cf. Lipton, 2004) in which that one of several candidate mental states is attributed that best fits the available evidence (Reisenzein, 2010). The *evidence* used to infer mental states includes diverse behavioral cues discussed in more detail below, but it also includes knowledge about possible eliciting events and the personality and history of the target person (if available). Everyday mental state detection is *theory-guided* because it makes use of an implicit theory of mind that specifies the typical connections between eliciting events, mental states, and behaviors (Heider, 1958; Malle, 2004).

Let us illustrate this with an example of everyday emotion inference. Imagine you tell your colleague Ann that you have met your common friend Oscar at a recent conference, whereupon Ann seems surprised. How did you come to think that Ann is surprised? You may have noticed that, in response to your communication, Ann showed one or more of the following behaviors: Her eyebrows raised; she interrupted her typing on the computer and turned to you; she said "What?" or even "This surprises me!" You may also have recalled, either before or after you informed Ann, that she had told you earlier that Oscar could not attend the conference because of an urgent family business, and you inferred from this that Oscar's appearance at the conference would be an unexpected, and hence surprising, event for Ann. Generalizing from this case, emotion inference in everyday life seems to proceed as follows: one observes that another person reacts in a particular way to a particular event, and one then uses one's implicit knowledge about the connections between eliciting events, mental states, and diverse behavioral indicators of mental states to infer that the other person probably experiences a particular emotion.

The observational approaches to the measurement of emotions described in this chapter can be regarded as scientific versions of the sketched process of everyday emotion inference (or part of it). Compared to everyday emotion inference, the scientific methods are typically restricted to a smaller set of emotions and a limited set of cues to emotion (most often facial expressions), and some are based on an explicit psychological theory of emotion (e.g., basic emotions theory; Ekman, 1972). However, the most important difference is that, whereas everyday observations of emotion-diagnostic behaviors and inferences to underlying emotions are usually made in an unsystematic, anecdotal manner, scientific observation is *systematic.* Systematic observation is marked by three main features (e.g., Huber, 1999):

1. Its explicit aim is the observation of a defined class of phenomena (objects, behaviors, activities, processes), and ideally the observer devotes his or her full attention to this task.
2. The observation is performed in a systematic or structured way, that is, it follows an observation plan or protocol—a set of predefined rules that specify *what* is to be observed, *when* it is to be observed, and *by whom*, how the observed behavior is to be *coded*, and related to this, which (if any) *interpretations* of the directly observable behaviors are to be made (Huber, 1999; see also Bakeman, 2000).
3. The quality of the obtained data is controlled by appropriate checks to ensure an acceptable quality level. The minimum quality requirement that a scientific observation method must meet—like any scientific measurement method—is a sufficient degree of reliability. *Reliability* refers to the precision of a measurement method; in the case of behavior observation, it is usually defined operationally as the degree of inter-observer agreement. For example, a facial expression coding system for detecting emotions is reliable to the degree that different observers (or groups of observers) infer the same emotions from the same facial expressions. However, as in the case of other psychological measurements, a behavior observation

method can be reliable but have low *validity*—it may not measure what it was designed to measure. To judge the validity of a behavioral emotion indicator, one would ideally like to compare it to a "gold standard"—a standard or criterion that unambiguously indicates the person's true emotional state. Although an entirely uncontested gold standard for emotion measurement does not exist, the most frequently used validity criterion is the self-report of emotion. There are two main reasons for this. First, self-reports of emotion can claim *epistemic priority:* the primary criterion for the presence of a particular emotion in a target person is that person's subjective experience, to which the experiencer, and only he or she, has direct access.[4] Second, self-reports of emotion have unmatched *specificity:* no other emotion indicator allows to distinguish as finely between different emotions (e.g., Reisenzein & Junge, 2012). Next to the self-report of emotion, the best validity criteria for behavioral emotion indicators are face-valid emotion induction methods (e.g., Reisenzein, Bördgen, Holtbernd, & Matz, 2006). For example, unexpected events are universally surprising, and certain stimuli are disgusting for nearly everyone.

Behavior Observation as a Form of Measurement

The systematic observation of behaviors, including the inference to underlying mental states, is a form of measurement in the broad sense of "measurement" introduced by Stevens (1946). According to this broad meaning, a method of emotion measurement is any method suited to determine the quality (e.g., joy, sadness, anger, fear) or intensity of emotions, provided that the results of the assessment are numerically coded (represented as numbers) in a consistent way. Before Stevens, the term "measurement" was restricted to methods for determining the amount or degree of a quantitative attribute, yielding numerical assignments on a metric scale (i.e., an interval or a higher scale level) (Michell, 1990). In the case of emotions, a candidate quantitative attribute does exist: It is the *intensity* of emotions, such as the degree of fear or the intensity of anger.

Behavioral indicators of emotion (e.g., facial expressions) are frequently only coded as present/absent by observers and are then typically used to diagnose the presence/absence of the underlying emotions. This amounts to nominal-scale level measurement in Stevens's (1946) sense. However, most behavioral indicators of emotion actually vary in intensity (e.g., smiles can range from just visible to highly intense), and it is typically assumed, in both common-sense and scientific psychology, that their intensity reflects that of the underlying emotion (e.g., amused people smile more strongly, the more amused they are). If so, codings of the intensity of emotion indicators by observers as well as observers' direct intensity ratings of the underlying emotions, should allow the measurement of emotion intensity on at least an ordinal scale level. Furthermore, if one assumes that the *probability* of an emotion-diagnostic behavior (e.g., brow-raising) increases with the intensity of the underlying emotion (e.g., surprise), one can also derive the intensity of emotion from the probability (estimated from the relative frequency) of behavior, and hence ultimately from measurements at the nominal scale level (e.g., Reisenzein, 2000). Again, the resulting intensity scale would be at least ordinal. Whether *metric measurement* (measurement at an interval or higher scale level) of emotion intensity is possible using behavior observation methods does not seem to have been systematically investigated. However, this question could in principle be answered by testing whether emotion intensity judgments by observers fulfill certain qualitative (ordinal) conditions called measurement axioms (Junge & Reisenzein, 2013; Krantz, Luce, Suppes, & Tversky, 1971).[5]

The preceding considerations refer to the behavioral measurement of emotion intensity at a particular time point or during a short time interval. To estimate the overall intensity of an emotion experienced during a longer time interval (e.g., the overall intensity of amusement felt while watching a humorous film clip), it has been proposed to use a composite of the frequency, intensity, and duration of diagnostic behavior (e.g., smiling) (e.g., Kring & Sloan, 2007).

Finally, note that although observational methods are often associated with nonexperimental (correlational) studies and with field research, and are indeed often used in these research contexts (Fernández-Dols & Crivelli, 2013), they are not restricted to them. On the contrary, being a method of *measurement,* the observation of emotion-related behavior can be used in both experimental and nonexperimental (correlational) research, and in laboratory as well as in field studies.

A Classification of Observational Methods of Emotion Measurement

Observational methods of emotion measurement can be classified according to several criteria, including: which emotions are covered, which kinds of cues to these emotions are considered (e.g., only facial expressions versus all available cues), how much inference is required, whether the rules of inference are made explicit or not, and relatedly, to which degree a method is based on an explicit emotion theory. Based mainly on the criteria of how much inference is required by an observational method and how explicit the rules of inference are, we distinguish between objective behavior coding systems, theory-based coding systems, and intuitive observer judgments of emotions. *Objective behavior coding systems* focus on the measurement of observable behaviors potentially diagnostic of emotions but make no inferences to underlying emotions (thus they are strictly speaking not measurements of *emotions* but only of *emotion-related behaviors*).[6] In contrast, the aim of the theory-based observation methods and of the intuitive observer approach is to infer emotions from behavior. *Theory-based behavior coding systems* are based, in part or completely, on scientific emotion theories, some of which also include assumptions about how emotions relate to particular behaviors, whereas *intuitive observer judgments of emotions* rely on observers' implicit beliefs, or their implicit folk-psychological theories, about the relationship between emotions and behavior. Before describing the different behavioral approaches to emotion measurement in more detail, we first discuss the nature and diagnostic value of those cues to emotion that are in principle available to human observers.

Observable Cues to Emotion

Observable cues to another person's current emotion comprise (1) the situation, by which we mean potentially emotion-eliciting events and the context in which they occur and (2) emotion-diagnostic behaviors of the target person. The behavioral indicators of emotion that are accessible to human observers can be classified according to whether they are intentional (i.e., nonverbal and verbal actions) or nonintentional (involuntary, although typically more or less controllable) behaviors.[7] Nonintentional behaviors potentially diagnostic of emotions comprise three main classes: (1) facial displays, (2) vocal displays, which include paralinguistic features of speech as well as nonlinguistic vocalizations and vocal bursts (see further on for an explanation), and (3) bodily displays, by which we mean postures, gestures, and body movements. In addition, (4) some of the involuntary physiological changes accompanying emotions, or side-effects of these bodily changes, can become visible to observers when they are intense (e.g., sweating, trembling).

Basic research on emotion-diagnostic behaviors has focused on nonintentional behaviors and among these, on facial and to a lesser extent, on vocal expressions (Harrigan, 2005; see also Calvo & D'Mello, 2010). The reason is that facial expressions are generally considered to be the best-discriminating nonverbal channel of emotion expression, followed by vocal and bodily expression. In the following section, we briefly review the main findings of this research.

Situational Information as a Cue to Emotion

Situational information potentially diagnostic of emotions comprises information about the nature of an emotion-eliciting event and the context in which it occurs. Such information can be highly predictive of the emotions induced by the event (e.g., Reisenzein & Hofmann, 1993). For example, Reisenzein and Hofmann found that naïve judges were able to infer with high accuracy (on average, 65% correct classifications) which of 23 emotions a target person experienced from brief descriptions of the eliciting situations. Evidence from this and other studies suggests, furthermore, that information about eliciting events is often available to observers in everyday life, either because they are personal witness to an eliciting event, or because they are informed about it by others including the target person (e.g., Rimé, 2009).

Intentional Actions as Cues to Emotion

Intentional actions potentially diagnostic of emotions comprise (1) gross motor actions presumably motivated by emotions (e.g., Heider, 1958; Weiner, 1995), such as hitting in the case of anger and helping in the case of pity and (2) verbal communications (speech acts) that convey information about emotions. While both kinds of cues can be quite useful for the inference of emotions in others, verbal communications are particularly diagnostic (Reisenzein & Junge, 2012). Emotion-diagnostic utterances comprise at least three different kinds: (1) Speech acts motivated by the emotion, such as an aggressive statement in anger or a comforting remark in pity. These communications can be regarded as verbal equivalents to nonverbal emotion-motivated actions, such as hitting (anger) or helping (pity); (2) Spontaneous self-reports of emotion (e.g., "I am so surprised"); (3) Descriptions of eliciting events (e.g., the target's—but also a third party's—report about an accident), which are substitutes for the direct observation of these events. Direct observations of the events that elicited a currently experienced emotion in another person are, in fact, not possible in many cases (Reisenzein & Junge, 2012)—for example, because the events have occurred in the past and are now only remembered.

Both nonverbal and verbal intentional actions have received comparatively little attention as cues to emotion in basic emotion research. However, the content of speech has been used in clinical diagnosis systems to infer emotions such as anxiety, hostility and depression, since at least the 1960s (Gottschalk & Gleser, 1969; see also, Gottschalk, 1995). It is also considered to some extent in the SPAFF system described further on

(Coan & Gottmann, 2007) as well as in recent research on automatic affect detection (Schuller, Batliner, Steidl, & Seppi, 2011).

Nonintentional Behaviors as Cues to Emotion: The Face

Research on the nonintentional—in particular, the facial—expression of emotions has been dominated by the theory of discrete basic emotions (e.g., Ekman, 1972, 1992; Izard & Dougherty, 1982). The adherents of this theory believe that a small subset of human emotions, considered to be biologically basic, are associated with distinct patterns of involuntary behaviors, in particular facial expressions (for more information, see the section on theory-based coding systems). According to Ekman's version of basic emotions theory, the emotions associated with distinct facial expressions include at minimum happiness, sadness, fear, anger, disgust, and surprise (e.g., Matsumoto & Ekman, 2008). The prototypical expressions of these six emotions are shown in Figure 29.1.

The main evidence for Ekman's theory stems from studies in which observers were presented with pictures of posed expressions of the basic emotions together with a list of their names and were asked to indicate which emotion is expressed by which facial expression (readers can try this for themselves with Figure 29.1). Using this method,

Figure 29.1 Prototypical facial expressions of six basic emotions according to Ekman (1972; see also Matsumoto & Ekman, 2008). From upper left to lower right: Happiness, sadness, anger, fear, disgust, and surprise. Photographs courtesy of Jörg Merten, Institute of Psychology, Saarland University.

average correct classification rates of more than 80% have been obtained in Western countries (e.g., Ekman et al., 1987; for reviews, see Elfenbein & Ambady, 2002; Nelson & Russell, 2013). However, accuracy is reduced if a free emotion labeling rather than a forced-choice method is used (Russell, 1994), and it also decreases with the distance to the Western culture (Nelson & Russell, 2013). Furthermore, these recognition studies show at best that if a target person shows a prototypical facial expression (which is actually not a statistical average but a high-intensity "ideal type"; Horstmann, 2002) then the correct emotion can be inferred with high accuracy. They do *not* show that people display the facial expression of a basic emotion whenever (and only) when they experience this emotion, and hence that the presence (and absence) of basic emotions can be reliably diagnosed from facial cues. Indeed, laboratory experiments (Reisenzein, Studtmann, & Horstmann, 2013) and naturalistic field studies (Fernández-Dols & Crivelli, 2013) of spontaneous emotional facial expressions suggest the opposite: Using self-reports of emotions or face-valid emotion induction methods as the criterion for the presence of emotions, these studies found that with the exception of amusement—which is usually not regarded as a basic emotion—only a minority of people who experience a discrete emotion show the facial expression presumably characteristic for it. Quite often (up to 90% in a series of studies on surprise by Reisenzein et al., 2006), no facial expression is shown at all, and if one occurs, it is typically only partial (i.e., only one or two components of the facial prototypes [Figure 29.1] are shown).

Nonintentional Behaviors as Cues to Emotion: The Voice

Speech can transmit information about emotions not only via the content of verbal messages (*what* is said, as described above) but also via *paralinguistic features* of vocal utterances (*how* something is said), such as pitch, voice intensity, and intonation. In addition, affective information can be conveyed via *nonlinguistic vocalizations*, such as breathing and laughter, and by what has been called *vocal bursts,* brief nonword utterances that arise between speech incidents and include shrieks, groans, and grunts as well as conventionalized expressions such as "wow!" (Schröder, 2003; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009).

Basic research on emotion recognition from the voice has focused on paralinguistic features of speech and has recently also looked at vocal bursts. The findings are similar to those obtained for facial expressions. Posed paralinguistic expressions of basic emotions (most often happiness, sadness, anger, and fear), typically obtained by asking actors to speak neutral or meaningless phrases in different intonations that express the different emotions, have yielded decoding accuracies corresponding to 70% correct in a forced-choice task with five response alternatives (Juslin & Laukka, 2003; see also Juslin & Scherer, 2005, 2008). This is somewhat less than the decoding accuracy obtained for posed facial expressions in the forced-choice paradigm (see previous discussion). Actor-posed vocal bursts for basic emotions (plus a few nonbasic emotions, such as amusement and relief) can be identified with similar accuracy (Sauter, Eisner, Calder, & Scott, 2010; Schröder, 2003; Simon-Thomas et al., 2009). However, analogous to the case of facial expression, studies on spontaneous vocal affect expression suggest that the "ideal-type" vocal expressions of basic emotions occur rarely in everyday speech (e.g., Cowie & Cornelius, 2003; Laukka, Neiberg, Forsell, Karlsson, & Elenius, 2011) and that correspondingly, the detection of emotions from natural vocal expressions is possible only to a limited extent. Nevertheless, beyond-chance detection of arousal level, the valence of the emotion (positive versus negative), and some specific emotions (such as irritation) from spontaneous speech seems possible (Laukka et al., 2011; Schuller et al., 2011).

Nonintentional Behaviors as Cues to Emotion: The Body

It has long been assumed that in contrast to facial and vocal expressions, bodily expressions (i.e., gestures, postures, and body movements) provide only information about the gross quality of affective states (e.g., positive vs. negative) and about the intensity of emotions but not about specific emotions (see Dael, Mortillaro, & Scherer, 2012a). However, a recent study in which professional actors were asked to portray 12 emotions in body actions and postures (Dael et al., 2012a) obtained evidence for discriminative patterns of bodily expression for at least three emotions (anger, amusement, and pleasure). For example, a characteristic expression for anger was the forward moving of the whole body, whereas pleasure was expressed by "head tilted up and averted" and an "asymmetrical arm action" (Dael et al., 2012a, p. 1090). Other research suggests that a number of nonstandard emotions and emotion-like states that are particularly important in academic contexts, such as interest, boredom, and confusion (e.g., Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010) can be detected with beyond-chance accuracy from body movements. Specifically, Mota and Picard (2003) measured posture patterns using the *Body Pressure Measurement System* (BMPS), a thin-film pressure pad with a rectangular grid of sensing elements that can be mounted on a variety of surfaces, such as the seat and back of a chair. They found that temporal transitions of posture patterns allowed to diagnose, with beyond-chance accuracy (75%), the interest level of children (as judged by teachers) while they performed a learning task on a computer. D'Mello and Graesser (2009) have made an attempt to extend this assessment method to the diagnosis of other achievement emotions.

OBSERVATION-BASED METHODS FOR EMOTION MEASUREMENT

Having discussed the nature and diagnostic value of the cues to emotion available to human observers, we now look more closely at the three main forms of observation-based methods for emotion measurement distinguished earlier: objective behavior coding systems, theory-based coding systems, and intuitive observer judgments of emotions.

Objective Behavior Coding Systems

FACS

The Facial Action Coding System (FACS; Ekman & Friesen, 1978) is an objective, anatomically based system for the measurement of (visible) facial behavior. Based on an anatomically based description of facial actions proposed by Hjortsjö (1969), FACS is considered the state-of-the-art instrument for the manual coding of facial movements. Its most recent version is FACS 2002 (Ekman, Friesen, & Hager, 2002; see also http://face-and-emotion.com/dataface/facs/manual/TitlePage.html). In FACS, facial expressions are coded in terms of *action units* (AUs), defined as the smallest possible

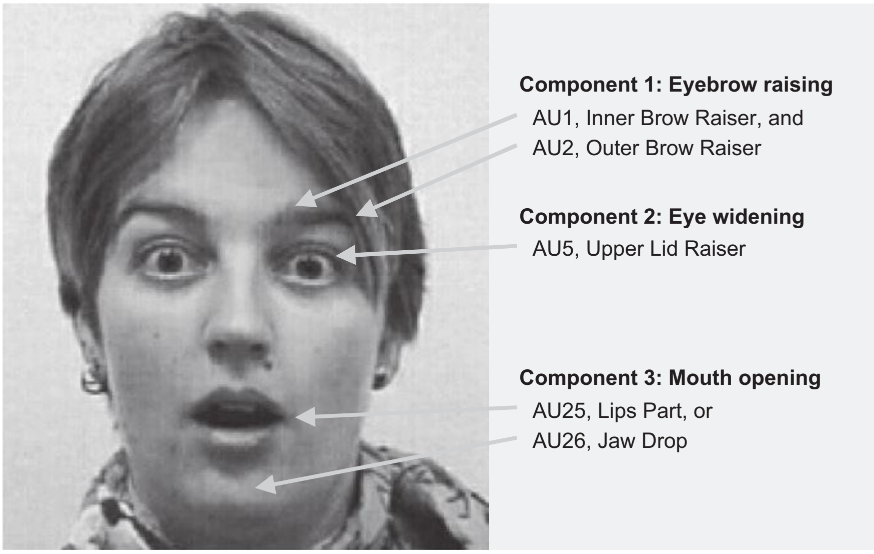

Figure 29.2 FACS coding of a prototypical surprise expression (codes according to Matsumoto & Ekman, 2008). Photograph courtesy of Jörg Merten, Institute of Psychology, Saarland University.

independent movements of facial muscles (Matsumoto & Ekman, 2008). These elementary movements can be regarded as the "phonemes" of facial expression (Littlewort et al., 2011). An overview of the FACS 2002 codes, including photographs of the action units, is given in Cohn, Ambadar, and Ekman (2007). As described there, FACS 2002 comprises 27 action units for the face (9 for the upper and 18 for the lower face), supplemented by 14 codes for head positions and movements, 9 for eye positions and movements, and several additional codes for other behaviors. Additionally, for some action units, FACS provides rules for scoring the intensity of the respective facial movements on a five-point scale. Figure 29.2 illustrates the FACS codes for a prototypical surprise face.

FACS coding can be performed comprehensively or selectively (Cohn et al., 2007). Comprehensive FACS coding considers the complete set of AUs, whereas selective FACS coding uses only a subset of AUs and ignores other facial movements. Comprehensive FACS coding is only possible for videos or still images because it is extremely time-consuming: According to Cohn et al. (2007), a well-trained FACS coder can take up to 100 minutes to code one minute of video data. However, comprehensive coding is not needed for many research purposes. If interest is restricted to Ekman's (1972) basic emotions, then EMFACS can be used (see below). If interest is still more narrowly restricted to a specific emotion (e.g., surprise, Reisenzein, 2000) or on specific facial movements (e.g., brow raising), then only the relevant subset of AUs need to be coded (see Figure 29.2 for the case of surprise). Alternatively, preliminary viewings of a set of videos may reveal that only a limited set of facial actions occur with sufficient frequency to warrant coding. Furthermore, FACS coding can be simplified by ignoring the intensities and the exact temporal onsets and offsets of AUs (which are also coded in comprehensive FACS coding)—that is, by coding only the occurrence and change of facial actions.

Cohn et al. (2007) summarize the findings of several studies on the reliability of well-trained FACS coders for 25 AUs. They found that reliability (expressed as Cohen's κ, the chance-corrected proportion of agreement) for presence/absence coding in a 0.5 sec tolerance window was good to excellent (about .70–.80) for nearly all AUs. As the tolerance window decreased in size, reliability decreased, but even with the smallest window (1/30th sec), 11 of 19 AUs had acceptable (> .60) reliabilities. High reliabilities have also been reported in studies using a strongly restricted set of AUs and coders trained only to recognize these AUs (e.g., Reisenzein, 2000). As to the validity of FACS, it is defined—because FACS is noninferential—as agreement with "what is really occurring on the face," as specified by an appropriate criterion such as an expert coding. Cohn et al. (2007) report that FACS codings agree well with several other validity criteria, such as the coding of instructed facial reactions and electromyographic (EMG) measurements of the involved facial muscles.

Objective Body Movement Coding and Voice Analysis Tools

An objective coding system analogous to FACS for the domain of *body action and posture*, the *Behavior Action and Posture coding system* (BAP), has been developed by Dael, Mortillaro, and Scherer (2012b). These authors also provide a review of other systems for coding body movement (see also, Harrigan, 2005). For *paralinguistic expressions of emotion,* a coding system using human observers analogous to FACS does not exist at present; however, as an alternative, objective acoustics-based methods are available that analyze speech waves by extracting parameters related to speech rate, voice intensity, fundamental frequency, and voice quality (e.g., Juslin & Scherer, 2005; Schuller et al., 2011). A much-used free software for the extraction of objective speech parameters is PRAAT (Boersma & Weenink, 2013).

Theory-Based Behavior Coding Systems

FACS is a tool for coding facial expressions objectively—that is, as facial movements—without making inferences to underlying emotions. In this sense, it is atheoretical (Matsumoto & Scherer, 2005). If the aim is to study whether or how particular emotions or other mental states are expressed in the face (e.g., Reisenzein et al., 2013), then no further inferences are in fact needed, and FACS is the method of choice. However, investigators of student and teacher emotions will typically not be interested in facial displays per se but in underlying emotional states. To get from facial movements to emotions, a set of inference rules is needed that connect facial behaviors (single AUs or combinations of AUs) to emotions. These rules are usually derived from a theory of emotion, such as basic emotions theory. Coding systems that contain such rules are therefore theory-based coding systems. Examples are EMFACS (Rosenberg & Ekman, 1984), the *Maximally Discriminative Facial Movement Coding System* (MAX) (Izard, 1979; Izard & Dougherty, 1982), and SPAFF (Coan & Gottman, 2007).

EMFACS (Emotion FACS)

EMFACS was developed on the basis of FACS with the aim to reduce scoring time when interest is only on emotion signals of the face. In EMFACS, only action units are coded that according to its authors are related to seven basic emotions (Ekman, 1972, 1992): the six already mentioned (happiness, sadness, anger, fear, disgust, and surprise; see Figure 29.1), plus contempt. In addition, the coding rules are simplified; in particular, whereas in standard FACS coding, the start and end time of each facial movement is coded, in EMFACS, the AUs are coded only at one time point (immediately before the point of maximum intensity). According to Ekman, Matsumoto, and Friesen (2005), coding time can be reduced to a tenth from FACS using EMFACS. Researchers considering using EMFACS should be aware, however, that the coding manual has not been published and is made available only to certified FACS coders.[8]

Although EMFACS could be regarded as a simplified version of FACS, the selection of the action units considered in EMFACS is grounded in basic emotions theory (Ekman, 1972). Furthermore, EMFACS users are typically interested in inferring basic emotions from the FACS codes and therefore use EMFACS together with a dictionary of assignment rules between FACS codes and emotions (e.g., Rosenberg, 2005). Descriptions of the assignment rules can be found, for example, in Matsumoto and Ekman (2008; see also, Ekman et al., 2002). These assignment rules, too, are based on basic emotions theory. For these reasons, EMFACS is here classified as a theory-based coding system.

The theoretical background of EMFACS, as said, is the theory of discrete basic emotions in the version proposed by Ekman (1972, 1992). According to basic emotions theory, the core of the emotion system consists of a set of discrete emotion mechanisms, each of which developed in evolution to solve a specific adaptive problem (e.g., the disgust mechanism developed to protect against poisoning by rotten food). If a basic emotion mechanism is evoked by suitable stimuli, it generates an emotion-specific pattern of responses including a specific feeling, a specific pattern of physiological reactions, and a characteristic facial expression (e.g., raising of the nose and upper lip in the case of disgust). According to Ekman, the main evolutionary function of the facial display is to communicate the emotional state to conspecifics. Although the facial expressions of the basic emotions can be deliberately suppressed or masked, as well as faked, basic emotions theory implies that spontaneous and uncontrolled displays reliably signal the presence of the corresponding basic emotions.

Assuming that the face-emotion assignment rules are unambiguous and are applied without error, the reliability of EMFACS codings depends only on the reliability of the corresponding FACS categories reported above and hence can be expected to be high. This is confirmed by the inter-coder agreements on EMFACS categories reported in several studies, which are high (e.g., Gottman & Levenson, 2002; Steimer-Krause, Krause, & Wagner, 1990). The validity of EMFACS can be gauged from the above-reported laboratory and field studies in which spontaneous expressions of basic emotions were related to self-reports of emotion or face-valid induction methods, particularly those in which EMFACS or FACS codings were made. These studies suggest that, with the exception of amusement, the validity of EMFACS is moderate to low: Expressed as a correlation, validities (not corrected for reliability) are approximately .65–.70 for amusement, < .50 for happiness, sadness, surprise, and disgust, and < .30 for anger and fear (Fernández-Dols et al., 2013; Reisenzein et al., 2013). However, note that in particular the data on anger and fear are sparse. Possible reasons for the moderate validity of facial expressions (and behavioral emotion indicators generally) as measures of emotion are discussed later.

Specific Affect Coding System (SPAFF)

The *Specific Affect Coding System* (SPAFF) was originally developed by Gottman and Krokoff (1989) for the systematic observation of "affective behavior" in marital conflict. It can be described as a theory-based coding system that (1) combines a set of explicit inference rules with an intuitive observer approach and (2) uses not only facial behaviors but also other nonverbal, as well as verbal cues including the content of utterances, to infer emotions and emotion-related action intentions (see further on). The development of SPAFF was motivated by Gottman's dissatisfaction with a previous, objective microanalytic coding system (CISS, Gottman, 1979) and similar coding systems such as FACS. According to Coan and Gottman (2007), these coding systems, which focus on physical features such as specific facial movements or gestures, often miss the forest for the trees because they are too discrete (they break up behavior in too small elements). To avoid this problem, SPAFF was devised to allow the direct coding of theoretical constructs (e.g., emotions). This is seen as a central advantage of SPAFF (Coan & Gottman, 2007, p. 267).

In its most recent version, SPAFF provides codes for 17 constructs referred to by Coan and Gottman (2007) as "positive affects" (affection, enthusiasm, humor, interest, and validation [meaning understanding and acceptance of the other]) and "negative affects" (anger, belligerence [a form of aggressive communication], contempt, criticism, defensiveness, disgust, domineering, fear/tension, sadness, stonewalling [unwillingness to listen or respond], threats, and whining). It is apparent that several of these categories (in particular validation, belligerence, criticism, defensiveness, and domineering) do not refer to emotions as typically conceived of by emotion researchers but are better described as interpersonal actions or interaction strategies, although they are probably partly motivated by emotions (e.g., defensiveness might be motivated by fear). These categories reflect SPAFF's origins as a tool to code marital interactions.

For each SPAFF construct, a hypothesized function in interpersonal encounters and a set of behavioral indicators are specified, and for a subset of the constructs, in addition a set of physical cues (facial expressions, postures, vocal features). Regarding the facial cues, SPAFF essentially incorporates EMFACS by interpreting particular combinations of FACS AUs as expressions of Ekman's basic emotions (sadness, anger, contempt, fear, disgust). However, other SPAFF categories are also linked to certain facial AUs (e.g., domineering). To illustrate, the SPAFF category *anger* is described as follows: (1) The function of anger is to "respond to perceived violations of the speaker's rights to autonomy and respect" (Coan & Gottman, 2007, p. 273); (2) Indicators of anger are frustration (a low-level form of an anger display marked by low-intensity facial expressions of anger and sometimes a lowering of the pitch and tempo of the voice), statements of being angry (e.g., "I am so angry!"), questions asked with angry affect and usually with sharp exhalations (as in "Why?!"), and commands (e.g., "Stop!") intended to stop a recent or ongoing violation of the speaker's autonomy and dignity; (3) Physical cues to anger are facial expressions of anger and changes in the voice (e.g., sudden increases in pitch, amplitude, and tempo). It is evident from this description that the indicators of SPAFF anger (and the same is true for the other categories) require considerable inference. For example, observers need to infer that frustration is present, that a question was asked with angry affect, or that a command was intended to stop a violation of the speaker's rights. To justify these inferences, Coan and Gottman (2007) refer to the so-called "cultural informants" approach to behavior observation, which assumes that experienced members of a culture are experts for the detection of emotional states from multiple nonverbal and verbal cues. This corresponds essentially to the intuitive observer approach to behavioral emotion measurement described in the next section. Hence, SPAFF combines a theory-based approach to the observational measurement of emotions with an intuitive observer approach.

SPAFF has been used in numerous studies to code affective behavior in interactions—mostly in couples, but also in children, their parents, and their peers (see Coan and Gottman, 2007; Jones, Carrère, & Gottman, 2005). Several studies found that adequate coding reliabilities can be obtained using SPAFF. For example, Carrère and Gottman (1999; see Jones et al., 2005) obtained reliabilities (Cronbach's alpha) > .70 for all SPAFF categories with the exception of contempt (.67), surprise (.56; this category was part of a previous version of SPAFF) and disgust (.37); the low reliabilities of disgust and surprise may, however, have been due to the fact that these categories occurred rarely. Similarly, Gottman and Levenson (2002) reported a chance-corrected proportion of agreement (Cohen's κ) of .75 for a 9-category version of SPAFF. Butler et al. (2003) reported an interrater reliability of $r = .90$ for a SPAFF-based index of positive emotions and .92 for negative emotions.

Although SPAFF codings have been shown to have predictive validity in being able to predict marital quality and divorce (e.g., Jones et al., 2005), data on the coherence of SPAFF emotion codings with self-reports of the same emotions are surprisingly scarce. In fact, we only found a single pertinent study (Geist & Gilbert, 1996). The authors found significant correlations to self-reported emotions for anger (.54) and sadness (.63) for wives, whereas for husbands, only the correlation for anger (.35) was significant.

Because SPAFF was explicitly designed for interaction situations, it should also be suited for use in academic contexts, although additional categories for emotions such as interest or boredom may have to be added. Finally, it may be noted that for eight SPAFF emotion categories, a behavior rating (rather than coding) system that allows nonexclusive intensity ratings has been developed (Johnson, 2002).

Facial Expression Coding System (FACES)

Developed by Kring, Smith, and Neale (1994), FACES, like EMFACS, is an exclusively face-based observational system for the measurement of emotions. Different from EMFACS and SPAFF, however, FACES does not refer to discrete basic emotions theory as its theoretical foundation but to the *dimensional approach* to emotion, specifically to pleasure-arousal theory (e.g., Russell, 1980, 2003; see also, Reisenzein, 1994). It is for this reason—the reference to pleasure-arousal theory—that we classify FACES as a theory-based observation method. However, apart from this and the fact that FACES uses some (modest) training of observers, FACES could also be classified as an intuitive observer judgment method for inferring valence (pleasure-displeasure) from facial expressions (Kring & Sloan, 2007) because it does not specify any face-emotion inference rules.

Pleasure-arousal theory assumes that emotional experiences, including basic emotions such as happiness, sadness, fear, and anger, or at least their feeling core (called "core affect" by Russell, 2003) consist of mixtures of more basic feelings—namely, feelings of pleasure or displeasure and of activation or deactivation. For example, the feeling core of anger is a mixture of displeasure and activation, whereas the feeling core of contentment is a mixture of pleasure and deactivation (see also, Reisenzein, 1994). As a consequence, pleasure-arousal theory rejects the assumption, made by some basic emotion theorists (e.g., McDougall, 1908; Oatley & Johnson-Laird, 1987), that basic emotions are characterized by unanalyzable feelings (see Reisenzein, 1995). In addition, dimensional emotion theorists reject the assumption that basic emotions are created by discrete affect programs that contain instructions for emotion-specific facial expressions (e.g., Barrett, 2006; Russell, 2003).

Approaches to emotion measurement based on pleasure-arousal theory are characterized by the attempt to measure the proposed components of emotion, either directly, by using items that ask participants to report the intensity of experienced pleasure-displeasure and arousal (e.g., Russell, Weiss, and Mendelsohn, 1989); or indirectly, by estimating the dimension values from measurements of specific affects (e.g., Smith, Vivian, & O'Leary, 1990). FACES uses the former approach; however, only the valence dimension (pleasure-displeasure) is assessed. Whenever a FACES rater detects a facial expression, defined as a change from a neutral face, he or she first makes a judgment about the valence of the expression (positive vs. negative). Next, the intensity of the facial expression (from 1 = low to 4 = high) is rated and finally its duration (in seconds).

Kring and Sloan (2007) present data on the reliability and validity of FACES. Concerning reliability, they report an average interrater agreement of *ICC* (intraclass correlation coefficient) = .86 in five studies, attesting to high reliability. As to the validity of FACES, (between-subject) correlations between observer ratings and self-reports of experienced pleasure/displeasure were found to depend on the nature of the emotion-eliciting event. Correlations were highest for amusing films ($r = .35$ to $r = .70$ for a composite of FACES frequency, intensity, and duration codings), moderate for happy ($r = .19$ to $r = .49$) and disgusting films ($r = .16$ to $r = .64$), and low for fearful films ($r = -.28$ to $r = .54$). This is similar to the validity of EMFACS inferences for Ekman's (1972) basic emotions, reported previously.

Intuitive Observer Judgments

An alternative to using formal behavior coding systems for the inference of mental states and traits is to use untrained raters (but possibly, and sometimes with a gain in validity [Sternglanz & DePaulo, 2004], people familiar with the target person, such as partners or staff). For example, ratings by peers are often used as observational measurements of personality traits (e.g., emotional stability, extraversion, conscientiousness) in personality psychology (Conolly, Kavanagh, & Viswesvaran, 2007). Applied to emotional states, the intuitive observer judgment method takes the intuitive observer approach already used as a component of SPAFF and FACES to its logical conclusion by dispensing completely with formal theory and relying entirely on observers' folk-psychological competence to infer emotions from behavior and context. The intuitive observer judgment method has been used in both basic research on the relation between emotions and facial expressions (e.g., Deckers, Kuhlhorst, & Freeland, 1987; see Reisenzein et al., 2013) and in more applied research, such as studies of interactions between couples (e.g., Waldinger, Schulz, Hauser, Allen, & Crowell, 2004).

Although the intuitive observer judgment method may at first sight appear to be a step backward when compared to more formal coding systems, such as SPAFF, it does have a number of advantages:

1. It is economical: Intuitive judges need not be specially trained, and the rating process takes no more time than comparable self-ratings do.
2. In principle, any emotion or emotion-related mental state can be judged by observers, including emotions such as interest and boredom, which are not considered in formal coding systems, such as EMFACS or SPAFF, but may be of particular interest to educational researchers (Pekrun et al., 2010). As a consequence, the intuitive

observer judgment method can be easily adjusted to the theoretical needs of the researcher.

3. The intuitive judgment method places no a priori restrictions on the cues used by observers to infer emotions, allowing them to use any available cue (facial, vocal, situational context, etc.) or cue combination. This maximally exploits the available information and best approximates the process of multicue emotion inference in everyday life.
4. Intuitive observer judgments avoid a potential problem of theory-based coding systems, which is the need to "reeducate" coders to define and recognize emotions in new ways that depart from their intuitive psychological understanding (Waldinger et al., 2004). In fact, Smith et al. (1990) reported that the attempt to train coders in the use of an observational coding system was partly unsuccessful, apparently because the coders' implicit theories of emotional expression were too deeply ingrained to change them in a reliable fashion.

Given these advantages of the intuitive observer judgment method, the decisive question becomes how it compares, in terms of reliability and validity, to the theory-based observational systems. Concerning reliability, the agreement between pairs of naïve judges is generally only moderate; however, reliability can be raised to adequate levels by pooling the judgments of several observers. In this way, individual biases of raters are minimized, and high agreement can be obtained (for details, see Rosenthal, 2005; see also, Schulz & Waldinger, 2005). For example, Waldinger et al. (2004) asked six judges to rate 18 emotions and emotion-related constructs (largely culled from the SPAFF) expressed in 30-second segments of interaction between couples. The reliability of the mean ratings of the six coders was on average r = .66, ranging from .27 (disgust) to .89 (humorous), with 14 of the 18 codings attaining reliabilities ≥ .60 and 11 ≥ .70. For combined indices of "hard emotions" (angry, annoyed, irritated, and aggravated) and "soft emotions" (hurt, sad, concerned, and disappointed), reliabilities around .90 were obtained by Sanford (2007). Concerning validity, Sanford (2007) reports correlations between observer ratings and self-ratings of .42 (wives) and .31 (husbands) for hard emotions, and .40 (wives) and .33 (husbands) for soft emotions. Although more research is needed, these findings suggest that the intuitive observer judgment method performs not much worse than the formal affect coding systems.

An interesting variant of intuitive observer ratings of emotion has been proposed by Levenson and Gottman (1983; see Ruef & Levenson, 2007). Observers use a dial to make near-continuous, moment-to-moment ratings of the intensity of an emotion (e.g., happiness or sadness) they perceive in the target person. These ratings can then be compared to analogous moment-to-moment self-ratings of emotion (e.g., Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005).

Reasons for and Implications of the Moderate Validity of Observational Methods

Whereas sufficient reliability can be attained for all described observational methods of emotion measurement, the validity of those methods that go beyond the observed behaviors to infer emotions (EMFACS, FACES, SPAFF, intuitive observer judgments) seems to have clear limits: With the exception of amusement (judged from facial behavior), validity as judged from agreements with self-reports typically does not exceed values of $r = .50$ for "basic emotions" (disgust, sadness, anger, fear) and is little better for pleasure-displeasure ratings. Possible reasons for the moderate coherence between emotions and their behavioral indicators have been most extensively discussed for facial expressions. The following explanations have been proposed:

1. Suboptimal designs used to estimate coherence—in particular, between-subjects rather than within-subjects designs (see, e.g., Reisenzein, 2000). Indeed, within-subjects correlations between emotion self-reports and facial expressions are usually higher than between-subjects correlations. However, with the exception of amusement, only moderate emotion-expression coherence is typically obtained even in within-subject designs (Reisenzein et al., 2013).
2. Measurement problems associated with self-reports of emotion, including the unwillingness or inability of people to accurately report the quality and intensity of their emotions (e.g., Rosenberg & Ekman, 1994). However, this is hardly the only reason: using conceptually identical self-report measures, comparatively high coherence between self-report and facial expression has been found for amusement (Reisenzein et al., 2013); conversely, low expression-experience coherence obtained for surprise was not found to increase when random measurement error in self-reports was reduced by an averaging method (Reisenzein, 2000).
3. Insufficient intensity of emotions. According to this hypothesis, emotions do not reveal themselves in facial expressions unless they exceed a threshold of intensity, which is often not reached in experimental and natural situations. Again, this is a possibility, but it explains at best part of the findings (Reisenzein et al., 2013).
4. Deliberate suppression or faking of emotional expressions (e.g., Rosenberg & Ekman, 1994). This is certainly a possibility; however, comparisons of facial expressions in social situations to those in solitary situations, where impression management should be less of an issue, suggest that expression control explains at best part of the findings (Reisenzein et al., 2013).
5. Emotions simply do not reveal themselves very clearly in facial behavior. At second thought, there may in fact be good evolutionary reasons for this: As noted by Fridlund (1994; see also Dawkins & Krebs, 1978; Russell, Bachorowski, & Fernández-Dols, 2003), the (involuntary or deliberate) truthful communication of emotions incurs potential costs to the sender, as it makes the sender more predictable and thus exploitable by others. In addition, by communicating his or her emotion to others, the sender may give away potentially useful information about the environment (e.g., that an unexpected event has occurred in the case of surprise) for free (Reisenzein & Junge, 2012). The truthful signaling of emotions to others is therefore a form of biological altruism that, like other altruistic behaviors, should have required special evolutionary conditions for its emergence. Possible evolutionary scenarios are kin selection, reciprocal altruism, group selection (Richerson & Boyd, 2005), and costly signaling. With the possible exception of costly signaling (see Dessalles, 2007), all of these scenarios require that emotions are not signaled indiscriminately but are revealed selectively to suitable targets—be it close kin, partners in a cooperative relationship, or members of a group with which the sender identifies. For other interaction partners, it would not be in the sender's interest to honestly communicate his or her emotions (and other mental states). Furthermore, it

> can be argued that for the purpose of selective emotion communication, language—human's main medium of communication—is in fact much better suited than non-verbal behavior (Reisenzein & Junge, 2012).

In view of the limited validity of behavioral cues to emotion, observational emotion researchers should try to use (1) multiple behavioral cues and (2) include additional information—in particular, about eliciting events, unless doing so interferes with the goals of the study (e.g., when the goal is to test whether a particular event induces a particular emotion). For an example of the proposed multiple-cue, theory-based inference of one emotion (surprise), see Reisenzein et al. (2006). Furthermore, note that although a moderate-validity measure of emotion is of limited use for the diagnosis of emotions on the level of *individuals* (which is typically the goal of emotion measurement in applied contexts, such as counseling or therapy), it can still be useful for research questions that can be answered by comparing *groups* (e.g., does marital counseling on average decrease interpersonal anger?). In addition, instead of using behavioral observations as moderately valid indicators of a target's true emotional state, they can be used as valid measures of how that state *is perceived by* others. This information can be highly valuable in itself because how a person's emotional state is perceived by others (e.g., a student's emotional state by teachers or the emotional well-being of a nursing home resident by staff [see Kolanowski, Hoffman, & Hofer, 2007]) is presumably more important for socially relevant consequences than the target's true emotional state. A parallel argument has been made for peer ratings of personality traits (see Conolly et al., 2007).

TECHNICAL ISSUES IN THE BEHAVIORAL OBSERVATION OF EMOTIONS

Online Versus Offline Coding

Coding of emotional behaviors can be performed online, that is while the behavior occurs; or retrospectively, using recordings of the behavior. Online coding by human observers presupposes that the coding is at all possible in real time and is therefore not an option for time-consuming coding systems such as FACS, unless a strongly restricted set of AUs is used (however, as described below, online FACS coding is now becoming possible using automatic coding systems). Online coding is an option, however, for FACES valence judgment and for intuitive observer judgments of emotion.

Although nowadays offline coding is typically preferred, it deserves to be pointed out that, when it is feasible, online coding actually has some advantages (see also, Bakeman, 2000). First, it requires no technical equipment but paper and pencil.[9] Second, online coding avoids potentially intrusive recording equipment such as visible cameras. Third, it allows observers to pick up cues to the target's emotion that are not available in video recordings (e.g., olfactory cues), or that can get lost in a video recording due to, for example, low speech volume, insufficient picture resolution, or a suboptimal camera angle.

The central disadvantage of online coding is the lack of a permanent behavior record and the increased observational possibilities that such a record affords. Therefore, behavior recordings should be made additionally even when online coding is possible and preferred. Although voice-only recordings (e.g., Schuller et al., 2011) and posture measurements (e.g., D'Mello & Graesser, 2009) can be an option in special cases, video recordings are the most useful and for this reason the most widely used behavior records. Coding of video recording has several advantages over online coding. First, prior to coding, the videos can be edited with video editing software in multiple useful ways. For example, to facilitate the coding procedure, critical sections of a teacher–student interaction can be cut and saved as separate video clips, or pasted together, and multiple video recordings of the same scene (e.g., one camera focusing on the teacher and another on the student) can be synchronized to be shown side by side. Second, the videos can be coded by multiple coders (which is particularly important for intuitive observer judgments to attain adequate reliability), and additional codings can be performed should new questions arise. Third, videos can be watched repeatedly and can be played back in slow motion or framewise to detect very brief or weak expressions.

Recording and Editing

Dinkelaker and Herrle (2009) recommend the use of digital camcorders because of their small size, easy operation, and high data compatibility. Videos can be recorded on Mini-DV (digital video) or high definition (HD) cassettes and later transferred to a personal computer or notebook for further processing. Alternatively, the videos can be stored directly on a computer. For the coding of facial expressions, care must be taken to obtain a good quality picture of the face (typically, a close-up of head and shoulders or head and upper body is sufficient). In educational settings, it is often useful if not imperative to use multiple cameras to optimally capture interactions between, for example, students and teacher. One camera focuses on the student and another on the teacher. A time-code (in millisecond or frame accuracy) that is visible in the video is very helpful and often indispensable. The time code is traditionally inserted into the camera video signal at recording time, using, for example, a Vertical Time Code (VITC) generator and a linked time code reader-plus-inserter. These hardware parts are offered as internal PC cards or as external boxes by several manufacturers at affordable prices. Alternatively, some professional video editing programs (e.g., FinalCut) as well as third-party software (e.g., TokiTC, www.tokitest.fr/english/tokitc.html) allow one to insert the time code later into the digitized video.

To facilitate the coding of interactions filmed by two cameras, it is useful to combine the two video streams in such a way that the behavior of the interaction partners (e.g., students and teachers) is shown synchronously side by side. For this, a special-effects generator with a "split screen" capability can be used, but there are also special video cards that allow the synchronous recording from several cameras. Alternatively, the two videos are recorded separately and are subsequently combined into a single video using video editing software, or they can simply be played back simultaneously side-by side (although this will usually need some manual adjustment to keep the videos in sync). If students are working on a computer, it is useful to synchronize the video track of their behavior with a video capture of the computer screen to be able to identify screen events as potential elicitors of expressive behavior. Educational researchers could also exploit the wide availability of computer labs at schools and universities to simultaneously record the behavior of multiple students: By attaching inexpensive webcams and microphones to the computers, they could obtain separate close-up recordings of every student in the lab.

Software for Coding

For the coding of the videos, a good video playing software is needed. Dinkelaker and Herrle (2009) recommend the freeware VLC Media Player. We have had good experiences with Zoom Player, which is also available in a Freeware version. Codes or ratings can be noted down on a sheet of paper with labeled columns or can be directly entered into a spreadsheet opened in a second window on the same or a separate monitor. This method actually works well for the coding of brief event-locked video recordings (e.g., student's facial reactions to teacher praise in a five-second window), particularly if the exact onset and offset times of the behaviors are not important. However, for the coding of more extended behavior streams and for the measurement of exact onsets and offsets, special video coding software can be a better choice. A variety of commercial and freeware/shareware programs are available for this purpose; some of them have been reviewed in Bakeman, Deckner, and Quera (2005). General-purpose commercial behavior observation software systems suited for the coding of emotional behaviors are INTERACT® (Mangold Software and Consulting, www.mangold.de/english/intoverview.htm) and OBSERVER® XT (Noldus company, www.noldus.com). Freeware/Shareware programs include multipurpose video annotation tools, such as ANVIL (Kipp, 2013, www.anvil-software.org) and ELAN (Lausberg, & Sloetjes, 2009; http://tla.mpi.nl/tools/tla-tools/elan/) and more specialized coding software, such as Etholog (Ottoni, 2000; www.ip.usp.br/docentes/ebottoni/EthoLog/ethohome.html) and ICODE (www-2.cs.cmu.edu/~face/index2.htm), which was specifically developed for FACS coding (see Cohn et al., 2007).

Automatic Coding and Affect Detection

Because of the time and effort required to code emotional behaviors, there have been, since the 1990s, attempts to develop computer programs that take over this task. Research on automatic coding of emotional behaviors and affect detection has increased greatly during the past decade, due in large part to the emergence of a new branch of computer science called *affective computing* that seeks to improve human-computer interaction by creating computer systems that are able to detect and appropriately respond to user emotions (Calvo, D'Mello, Gratch, & Kappas, 2014; Picard, 1997). A recent overview of automatic affect detection is provided by Calvo and D'Mello (2010).

As might be expected, a focus of this research is the development of programs that allow to detect emotion in the face (e.g., Cohn & Kanade, 2007; Pantic & Bartlett, 2007; Zeng, Pantic, Roisman, & Huang, 2009). One approach is to develop programs that first detect FACS action units in videos, from which emotions (or other mental states) can then be inferred using theoretical or empirically determined assignment rules. A second approach to automatic facial emotion recognition attempts to infer emotions directly from low-level image features (see Pantic & Bartlett, 2007). An example system that allows online coding of both FACS action units and basic emotions is the *Computer Expression Recognition Toolbox* (CERT) (Littlewort et al., 2011). Figure 29.3 illustrates the operation of CERT for the recognition of FACS AUs (for details, see Littlewort et al., 2011).

Although there has been remarkable progress in automatic FACS scoring during the past years, the performance of current automatic AU detection systems does not yet quite match that of human coders (Calvo & D'Mello, 2010). Furthermore, most automatic facial scoring systems are research prototypes. At least two real-time automatic facial

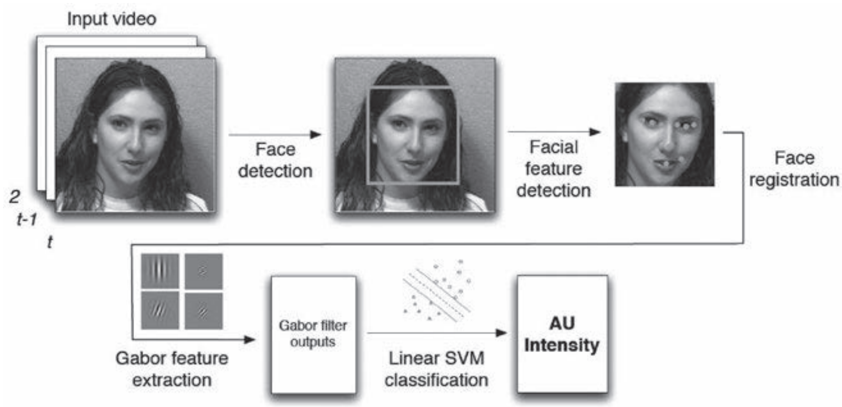

Figure 29.3 Processing pipeline of the Computer Expression Recognition Toolbox (CERT) from video to AU intensity estimates. From Littlewort et al. (2011). Courtesy of Gwen Littlewort, Machine Perception Laboratory, University of California, San Diego.

coding systems, however, are publicly available: FaceReader™ (D'Arcey, Johnson, & Ennis, 2012) and FACET™ (http://imotionsglobal.com/software/add-on-modules/attention-tool-facet-module-facial-action-coding-system-facs/), which is a commercial system based on the CERT technology (Littlewort et al., 2011).

During the past 10 years, the voice has also become a favorite target of automated affect detection systems. Several audio-based automatic emotion detection systems are reviewed by Zeng et al. (2009; see also, Calvo & D'Mello, 2010; Schuller et al., 2011). Although many of these systems focus on basic emotions (Ekman, 1992), efforts have also been made to detect other affective states, such as frustration (e.g., Laukka et al., 2011).

Automatic systems have also been developed for the detection of affect from posture (e.g., D'Mello & Graesser, 2009) and from physiological reactions (see Calvo & D'Mello, 2010). Furthermore, there is a trend to develop affect detection systems that fuse the information from several channels (Calvo & D'Mello, 2010). Although these latter systems are at present still research prototypes, educational researchers might team up with computer scientists working in the field of affective computing (see, e.g., http://emotion-research.net/) to mutual profit.

We consider it possible that within the next 10 years, automatic affect detection tools systems will reach the stage where they can compete with human observers. To achieve this, these systems will probably need to combine accurate sensing of multiple behavioral cues, including the content of speech, with knowledge about the context and the target's personality and history, and with an elaborated theory of mind component that specifies the links between behavioral cues and emotions and is possibly adjusted to the specific targets during training sessions (Reisenzein, 2010). It is conceivable that automatic affect detection systems will eventually even outperform human observers, either because they include signals (e.g., from subtle physiological changes) not available to human observers or because they use emotion inference algorithms that outperform those implicitly used by humans. In any case, the currently used observational methods of emotion measurement, described in the core section of this chapter, may soon become replaced by computer-based systems. Finally, even though there may be evolutionary limits to what can be detected about emotional states from the observation of behaviors, measurements of brain activity are not so limited: Because on the neurophysiological level, emotional states are brain states, measurements of brain activity could eventually provide precise external measurements of both the quality and intensity of experienced emotions. For an example of recent research in this area, see Wagner, Atlas, Lindquist, Roy, Woo, and Kross (2013).

NOTES

1. Empirical methods of knowledge acquisition are traditionally juxtaposed to rational methods (methods based on reasoning; e.g., Musgrave, 1993), which can be defined as all valid methods of drawing inferences from existing knowledge (e.g., deduction, induction, inference to the best explanation). Rational methods play an indispensable role in all sciences, and some sciences (e.g., logics and mathematics) use them exclusively. Inference also plays an essential role in many psychological observation methods, as illustrated in this chapter by the theory-based and intuitive observational approaches to emotion measurement.
2. This is not meant to be a precise definition. In fact, we believe with others (e.g., Cranach & Frenz, 1969) that a sharp demarcation of behavior observation from other methods of external observation is not possible. But neither is it needed.
3. Broadly understood, mental state detection also includes the inference of personality traits (Schneider et al., 1979). Many of these traits actually are, or involve, dispositions to have particular emotions (Reisenzein & Weber, 2009).
4. For a recent critique of this assumption, see Jäger (2009). Jäger argues that, as a matter of fact, we often suffer from affective ignorance, as a consequence of which observation-based ascriptions of emotions should often be credited with more rather than less authority than corresponding self-ascriptions. Although this viewpoint is a minority position among today's emotion researchers, it should be acknowledged that self-reports of emotions are subject to a number of possible biases, including reactivity (observing one's emotional state may alter this state), interindividual differences in the meanings of the emotion concepts used to report one's feelings, and unwillingness to report one's emotions (see, e.g., Mauss & Robinson, 2009).
5. Self-reports of emotion intensity on rating scales (e.g., "How happy are you right now on a scale from 0 = not at all to 10 = extremely") are thought to lie somewhere in between the ordinal and interval scale level (Krantz, Luce, Suppes, & Tversky, 1971); an interval scale level can, however, be reached with alternative self-report methods based on comparative judgments (Junge & Reisenzein, 2013). The same may be true for observer judgments of emotion intensity.
6. Note, however, that some emotion theorists consider emotion-related behaviors (at least certain involuntary behaviors) to be part of emotions, which they conceptualize as mental-behavioral syndromes (e.g., Lazarus, 1991).
7. Although the theoretical distinction between intentional and nonintentional emotion-related behaviors is widely accepted, the following classification of some kinds of emotional behavior as intentional versus unintentional could be questioned. For example, it could be argued that most gestures and some vocal bursts are intentional, whereas goal-directed actions that have become habitual should be considered unintentional. Furthermore, because most nonintentional behaviors can be deliberately simulated, the classification of *concrete instances* of emotional behaviors (e.g., Ann's smile at the joke Bill told her) as intentional versus unintentional is always more or less tentative.
8. Certified FACS coders have passed a test issued by Ekman's research group (see Ekman et al., 2002, and http://face-and-emotion.com/dataface/facs/FFT_Proc.html). It takes around 100 hours of working through the FACS manual to become competent enough to take the test (Ekman et al., 2005). Alternatively, one can participate in a FACS training course offered by several FACS researchers (Cohn et al., 2007).
9. Nevertheless, a hand-held electronic device is recommended for entering the codes because it prevents data entry errors when transferring the data from the sheets and also allows automatic storing of the time. For a review of hand-held data entry devices, see Adiguzel, Vannest, and Zellner (2009).

REFERENCES

Adiguzel, T., Vannest, K. J., & Zellner, R. D. (2009). The use and efficacy of handheld computers for school-based data collection: A literature review. *Computers in the Schools, 26,*187–206.

Bakeman, R. (2000). Behavioral observations and coding. In H. T. Reis & C. K. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 138–159). New York, NY: Cambridge University Press.

Bakeman, R. Deckner, D. F., & Quera, V. (2005). Analysis of behavioral streams. In D. M. Teti (Ed.), *Handbook of research methods in developmental science* (pp. 394–420). Oxford, United Kingdom: Blackwell Publishers.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science, 1,* 28–58.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.4.40) [Computer program]. Retrieved Feb. 5, 2013 from www.praat.org/

Brentano, F. (1874). *Psychologie vom empirischen Standpunkte* [*Psychology from the empirical standpoint*]. Leipzig, Germany: Duncker & Humblot.

Butler, E. A., Egloff, B., Wilhelm, F. H., Smith, N. C., Erickson, E. A., & Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion, 3,* 48–67.

Calvo, R. A., & D'Mello, S. K. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE *Transactions on Affective Computing, 1,* 18–37.

Calvo, R. A., D'Mello, S. K., Gratch, J., & Kappas, A. (Eds.). (2014). *Handbook of affective computing.* Oxford, United Kingdom: Oxford University Press.

Carrère, S., & Gottman, J. M. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family Process, 38,* 293–301.

Coan, J. A., & Gottman, J. M. (2007). The Specific Affect coding system (SPAFF). In J. A. Coan & J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 106–123). New York, NY: Oxford University Press.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan & J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 203–221). New York, NY: Oxford University Press.

Cohn, J. F., & Kanade, T. (2007). Automated facial image analysis for measurement of emotion expression. In J. A. Coan & J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 222–238). New York, NY: Oxford University Press.

Conolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment, 15,* 110–117.

Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40,* 5–32.

Cranach, M. V., & Frenz, H.-G. (1969). Systematische Beobachtung [Systematic observation]. In C. F. Graumann (Ed.), *Handbuch der Psychologie* [Handbook of psychology] (Vol 7.1; pp. 269–331). Göttingen, Germany: Hogrefe.

D'Arcey, T., Johnson, M., & Ennis, M. (2012). Assessing the validity of FaceReader using facial electromyography. *Proceedings of APS 24th annual meeting.* Retrieved from www.darcey.us/pdf/facereader.pdf

D'Mello, S., & Graesser, A. (2009). Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence, 23,* 123–150.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012a). Emotion expression in body action and posture. *Emotion, 12,* 1085–1101.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012b). The Body Action and Posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior, 36,* 97–121.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology* (pp. 282–309). Oxford, United Kingdom: Blackwell.

Deckers, L., Kuhlhorst L., & Freeland, L. (1987). The effects of spontaneous and voluntary facial reactions on surprise and humor. *Motivation and Emotion, 11,* 403–412.

Dessalles, J.-L. (2007). *Why we talk: The evolutionary origins of language.* Oxford, United Kingdom: Oxford University Press.

Dinkelaker, J., & Herrle, M. (2009). *Erziehungswissenschaftliche Videographie: Eine Einführung* [*Educational videography: An introduction*]. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation* (Vol. 19, pp. 207–283). Lincoln, NE: University of Nebraska Press.

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

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. (2002). *Facial Action Coding System.* Salt Lake City, UT: Research Nexus.

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

Ekman, P., Matsumoto, D., & Friesen, W. (2005). Facial expression in affective disorders. In P. Ekman & E. L. Rosenberg (Eds.), *What the face reveals: Basic and applied studies of. spontaneous expression using the Facial Action Coding System* (FACS) (pp. 429–439). New York, NY: Oxford University Press.

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

Fernández-Dols, J. M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review, 5,* 24–29.

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

Geist, R. L., & Gilbert, D. G. (1996). Correlates of expressed and felt emotion during marital conflict: satisfaction, personality, process, and outcome. *Personality and Individual Differences, 21,* 49–60.

Gottman, J. M. (1979). *Marital interaction: Experimental investigations.* New York, NY: Academic Press.

Gottman, J. M., & Krokoff, L. J. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57,* 47–52.

Gottman, J., & Levenson, R.W. (2002). A two-factor model for predicting when a couple will divorce: Exploratory analyses using 14-year longitudinal data. *Family Process, 41,* 83–96.

Gottschalk, L. A. (1995). *Content analysis of verbal behavior. New findings and clinical applications.* Hillsdale, NJ: Lawrence Erlbaum.

Gottschalk, L. A., & Gleser, G. C. (1969). *The measurement of psychological states through the content analysis of verbal behavior.* Los Angeles, CA: University of California Press.

Harrigan, J. A. (2005). Proxemics, kinesics, and gaze. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *New handbook of methods in nonverbal behavior research* (pp. 137–198). New York, NY: Oxford University Press.

Heider, F. (1958). *The psychology of interpersonal relations.* New York, NY: Wiley.

Hjortsjö, C. H. (1969). *Man's face and mimic language.* Lund, Sweden: Student-Litteratur.

Horstmann, G. (2002). Facial expressions of emotion: Does the prototype represent central tendency, frequency of instantiation, or an ideal? *Emotion, 2,* 297–305.

Huber, O. (1999). Beobachtung [Observation]. In E. Roth, H. Holling, & K. Heidenreich (Eds.), *Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis* [*Social science methods: Textbook and handbook for research and practice*] (pp. 126–145). München, Germany: Oldenburg.

Immordino-Yang, M. H., & Christodoulou, J. A. (2014). Neuroscientific contributions to understanding and measuring emotions in educational contexts. In R. Pekrun & L. Linnenbrink-Garcia (Eds.) *International handbook of emotions in education*. New York, NY: Routledge.

Izard, C. E. (1979). *The Maximally Discriminative Facial Movement Coding System (MAX).* Newark, DE: University of Delaware Office of Instructional Technology.

Izard, C., & Dougherty, L. (1982). Two complementary systems for measuring facial expressions in infants and children. In C. Izard (Ed.), *Measuring emotions in infants and children* (pp. 97–126). New York, NY: Cambridge University Press.

Jäger, C. (2009). Affective ignorance. *Erkenntnis*, *71*, 123–139.

Johnson, M. D. (2002). The observation of specific affect in marital interactions: Psychometric properties of a coding system and a rating system. *Psychological Assessment, 14,* 423–438.

Jones, S., Carrère, S., & Gottman, J. M. (2005). Specific Affect Coding System. In V. L. Manusov (Ed.), *The sourcebook of nonverbal measures: Going beyond words* (pp. 163–172). Mahwah, NJ: Erlbaum.

Junge, M., & Reisenzein, R. (2013). Indirect scaling methods for testing quantitative emotion theories. *Cognition and Emotion, 27*, 1247–1275.

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. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *New handbook of methods in nonverbal behavior research* (pp. 65–136). New York, NY: Oxford University Press.

Juslin, P. N., & Scherer, K. R. (2008). Speech emotion analysis. Received from www.scholarpedia.org/article/Speech_emotion_analysis

Kipp, M. (2013). ANVIL: A universal video research tool. In J. Durand, U. Gut, & G. Kristofferson (Eds.), *Handbook of corpus phonology.* Oxford, United Kingdom: Oxford University Press.

Kolanowski, A., Hoffman, L., & Hofer, S. M. (2007). Concordance of self-report and informant assessment of emotional well-being in nursing home residents with dementia. *Journal of Gerontology: Psychological Sciences, 62B,* 20–27.

Krantz, D., Luce, R., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Vol. 1: Additive and polynomial representations.* New York, NY: Academic Press.

Kreibig, S. D., & Gendolla, G.H.E. (2014). Autonomic nervous system measurement of emotion in education and achievement settings. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 625–642). New York, NY: Taylor & Francis.

Kring, A. M., & Sloan, D. M. (2007). The facial expression coding system (FACES): Development, Validation and Utility. *Psychological Assessment, 19,* 210–224.

Kring, A. M., Smith, D. A., & Neale, J. M. (1994). Individual differences in dispositional expressiveness: The development and validation of the Emotional Expressivity Scale. *Journal of Personality and Social Psychology, 66,* 934–949.

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

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers, 41,* 841–849.

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

Levenson, R. W., & Gottmann, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and social Psychology, 45,* 587–597.

Littlewort, G., Whitehill, J., Wu, T., Fasel, I. R., Frank, M., Movellan, J. R., & Bartlett, M. S. (2011). The computer expression recognition toolbox (CERT). *Proceedings of the 9th IEEE Conference on Automatic Face and Gesture Recognition* (pp. 298–305). Santa Barbara, CA.

Lipton, P. (2004). *Inference to the best explanation.* London, United Kingdom: Routledge.

Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction.* Cambridge, MA: MIT Press.

Matsumoto, D., & Ekman, P. (2008). Facial expression analysis. *Scholarpedia, 3*(5), 4237.

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5,* 175–190.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23,* 209–237.

McDougall, W. (1908/1960). *An introduction to social psychology.* London, United Kingdom: Methuen.

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Erlbaum.

Mota, S., & Picard, R. (2003, June). Automated posture analysis for detecting learner's interest level. *Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction,* CVPR HCI.

Musgrave, A. (1993). *Common Sense, science and scepticism: A historical introduction to the theory of knowledge.* Cambridge, United Kingdom: Cambridge University Press.

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*, 8–15.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion, 1,* 29–50.

Ottoni, E. B. (2000). EthoLog 2.2—A tool for the transcription and timing of behavior observation sessions. *Behavior Research Methods, Instruments, and Computers, 32,* 446–449.

Pantic, M., & Bartlett, M. S. (2007). Machine analysis of facial expressions. In K. Delac & M. Grgic (Eds.), *Face recognition* (pp. 377–416). Vienna, Austria: I-Tech Education and Publishing.

Pekrun, R., & Bühner, M. (2014). Self-report measures of academic emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education*. New York, NY: Taylor & Francis.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102,* 531–549.

Picard, R. W. (1997). *Affective computing.* Cambridge, MA: MIT Press.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67,* 525–539.

Reisenzein, R. (1995). On Oatley and Johnson-Laird's theory of emotions and hierarchical structures in the affective lexicon. *Cognition and Emotion, 9,* 383–416.

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion, 14,* 1–38.

Reisenzein, R. (2010). Broadening the scope of affect detection research. *IEEE Transactions on Affective Computing, 1,* 42–45.

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology, 91,* 295–315.

Reisenzein, R., & Hofmann, T. (1993). Discriminating emotions from appraisal-relevant situational information: Baseline data for structural models of cognitive appraisals. *Cognition and Emotion, 7,* 271–293.

Reisenzein, R., & Junge, M. (2012). Language and emotion from the perspective of the computational belief-desire theory of emotion. In P. A. Wilson (Ed.), *Dynamicity in emotion concepts.* Lodz Studies in Language 27 (pp. 37–59). Frankfurt am Main, Germany: Peter Lang.

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review, 5,* 16–23.

Reisenzein, R., & Weber, H. (2009). Personality and emotion. In P. J. Corr & G. Matthews (Eds.), *Cambridge handbook of personality psychology* (pp. 54–71). Oxford, United Kingdom: Oxford University Press.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution.* Chicago, IL: University of Chicago Press.

Rimé, B. (2009). Emotions elicit the social sharing of emotion: Theory and empirical review. *Emotion Review, 1,* 60–85.

Rosenberg, E. (2005). The study of spontaneous facial expressions in psychology. In P. Ekman & E. L. Rosenberg (Eds.), *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System* (FACS) (pp. 3–18). New York, NY: Oxford University Press.

Rosenberg, E. L., & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition and Emotion, 8,* 201–229.

Rosenthal, R. (2005). Conducting judgment studies: Some methodological issues. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *New handbook of methods in nonverbal behavior research* (pp. 199–234). New York, NY: Oxford University Press.

Ruef, A. M., & Levenson, R. W. (2007). Continuous measurement of emotion. In J. A. Coan & J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 286–297). New York, NY: Oxford University Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178.

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

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54,* 329–349.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57,* 493–502.

Sanford, K. (2007). The couples emotion rating form: Psychometric properties and theoretical associations. *Psychological Assessment, 19,* 411–421.

Sauter, D., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology, 63,* 2251–2272.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person Perception.* Reading, MA: Addison-Wesley.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40,* 99–116.

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognizing realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication, 53,* 1062–1087.

Schulz, M. S., & Waldinger, R. J. (2005). The value of pooling "naïve" expertise. *American Psychologist, 60,* 656–662.

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, D. A., Vivian, D., & O'Leary, K. D. (1990). Longitudinal prediction of marital discord from premarital expressions of affect. *Journal of Consulting and Clinical Psychology, 58,* 790–798.

Steimer-Krause, E., Krause, R., & Wagner, G. (1990). Interaction regulations used by schizophrenic and psychosomatic patients: studies on facial behavior in dyadic interactions. *Psychiatry, 53,* 209–228.

Sternglanz, R. W., & DePaulo, B. M. (2004). Reading nonverbal cues to emotions: The advantages and liabilities of relationship closeness. *Journal of Nonverbal Behavior, 28,* 245–266.

Stevens, S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-Based neurologic signature of physical pain. *New England Journal of Medicine, 368,* 1388–1397.

Waldinger, R. J., Schulz, M. S., Hauser, S. T., Allen, J. P., & Crowell, J. A. (2004). Reading other's emotions: The role of intuitive judgments in predicting marital satisfaction, quality and stability. *Journal of Family Psychology, 18,* 58–71.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct.* New York, NY: Guilford.

Wright, H. F. (1960). Observational child study. In P. H. Mussen (Ed.), *Handbook of research methods in child development* (pp. 71–139). New York, NY: Wiley.

Wundt, W. (1896). *Grundriss der Psychologie* [Outlines of psychology]. Leipzig, Germany: Engelmann.

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