Jeffrey F. Cohn
Zara Ambadar
Paul Ekman

13

Observer-Based Measurement of Facial Expression With the Facial Action Coding System

Facial expression has been a focus of emotion research for over a hundred years (Darwin, 1872/1998). It is central to several leading theories of emotion (Ekman, 1992; Izard, 1977; Tomkins, 1962) and has been the focus of at times heated debate about issues in emotion science (Ekman, 1973, 1993; Fridlund, 1992; Russell, 1994). Facial expression figures prominently in research on almost every aspect of emotion, including psychophysiology (Levenson, Ekman, & Friesen, 1990), neural bases (Calder et al., 1996; Davidson, Ekman, Saron, Senulis, & Friesen, 1990), development (Malatesta, Culver, Tesman, & Shephard, 1989; Matias & Cohn, 1993), perception (Ambadar, Schooler, & Cohn, 2005), social processes (Hatfield, Cacioppo, & Rapson, 1992; Hess & Kirouac, 2000), and emotion disorder (Kaiser, 2002; Sloan, Straussa, Quirka, & Sajatovic, 1997), to name a few.

Because of its importance to the study of emotion, a number of observer-based systems of facial expression measurement have been developed (Ekman & Friesen, 1978, 1982; Ekman, Friesen, & Tomkins, 1971; Izard, 1979, 1983; Izard & Dougherty, 1981; Kring & Sloan, 1991; Tronick, Als, & Brazelton, 1980). Of these various systems for describing facial expression, the Facial Action Coding System (FACS; Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002) is the most comprehensive, psychometrically rigorous, and widely used (Cohn & Ekman, 2005; Ekman & Rosenberg, 2005). Using FACS and viewing video-recorded facial behavior at frame rate and slow motion, coders can manually code nearly all possible facial expressions, which are decomposed into action units (AUs). Action units, with some qualifications, are the smallest visually discriminable facial movements. By comparison, other systems are less thorough (Malatesta et al., 1989), fail to differentiate between some anatomically distinct movements (Oster, Hegley, & Nagel, 1992), consider movements that are not anatomically distinct as separable (Oster et al., 1992), and often assume a one-to-one mapping between facial expression and emotion (for a review of these systems, see Cohn & Ekman, in press).

Unlike systems that use emotion labels to describe expression, FACS explicitly distinguishes between facial actions and inferences about what they mean. FACS itself is descriptive and includes no emotion-specified descriptors. Hypotheses and inferences about the emotional meaning of facial actions are extrinsic to FACS. If one wishes to make emotion-based inferences from FACS codes, a variety of related resources exist. These include the FACS Investigators' Guide (Ekman et al., 2002), the FACS interpretive database (Ekman, Rosenberg, & Hager, 1998), and a large body of empirical research.(Ekman & Rosenberg, 2005). These resources suggest combination rules for defining emotion-specified expressions from FACS action units, but this inferential step remains extrinsic to FACS. Because of its descriptive power, FACS is regarded by many as the standard measure for facial behavior and is used widely in diverse fields. Beyond emotion science, these include facial neuromuscular disorders (Van Swearingen & Cohn, 2005), neuroscience (Bruce & Young, 1998; Rinn, 1984, 1991), computer vision (Bartlett, Ekman, Hager, & Sejnowski, 1999; Cohn, Zlochower, Lien, & Kanade, 1999; Pantic & Rothkrantz, 2000; Tian, Cohn, & Kanade, 2005), computer graphics and animation (Breidt, Wallraven, Cunningham, & Buelthoff, 2003; Parke & Waters, 1996), and face encoding for digital signal processing (International Organization for Standardization, 2002; Tao, Chen, Wu, & Huang, 1999).

In this chapter, we discuss the conceptual basis for FACS, the numerical listing of discrete facial movements identified by the system, the evaluative psychometrics of the system, and the recommended training requirements. We also include information on how to obtain software for computer-assisted FACS coding.

FACS has progressed through three versions: the initial version (FACS 1978), a document-based update (FACS 1992), and a new edition (FACS 2002), which includes improvements in scoring criteria and in didactic materials, extensive use of hyperlinked cross-referenced text, and embedded video links in the CD version. Throughout this chapter, we use publication date when referring to a specific version of FACS.

Conceptual Basis

Sign Versus Message Judgment

Ekman and Friesen (Ekman, 1964, 1965; Ekman & Friesen, 1969) distinguished two conceptual approaches to studying facial behavior, namely, measuring judgments about one or another message and measuring the sign vehicles that convey the message. In message judgment, the observer's task is to make *inferences* about something underlying the facial behavior—emotion, mood, traits, attitudes, personality, and the like; for this reason observers typically are referred to as "judges" or "raters." In measuring sign vehicles, observers *describe* the surface of behavior; they count how many times the face moves a certain way, or how long a movement lasts, or whether it was a movement of the *frontalis* or *corrugator* muscle. As an example, on seeing a smiling face, an observer with a judgment-based approach would make judgments such as "happy," whereas an observer with a sign-based approach would code the face as having an upward, oblique movement of the lip corners. Observers with a sign-based approach are supposed to function like machines and typically are referred to as "coders."

Though message- and sign-based approaches can sometimes answer the same questions, they can also answer different questions, for they focus on different phenomena. Message judgment research is not typically focused on the face. The face is but an input. The focus is on the person observing the face and/or on the message obtained. Questions have to do with whether a difference is detectable or accurate, whether there are individual differences among raters, reflecting skill, gender, or personality, and whether messages obtained are best represented as dimensions or categories.

Facial sign vehicles are measured when the focus is on unearthing something fairly specific about facial behavior itself, not about the perception of the face. It is the only method that can be used to answer such questions as:

1. To what extent is the facial activity shown by newborns and infants systematic, not random, and which particular actions first show such systematic organization? To answer this question, facial behavior shown during samples taken at different developmental points or in different situational contexts can be measured. Then the probabilities of particular co-occurrences and sequential patterns of facial actions can be evaluated (Cohn & Tronick, 1983; Oster & Ekman, 1978).
2. Which particular facial actions are employed to signal emphasis in conversation? Facial actions that co-occur with verbal or vocal emphasis must be measured to determine whether there are any actions that consistently accompany any emphasis (Ekman, 1980).
3. Is there a difference in the smile during enjoyment as compared with a discomfort smile? The particular facial actions evident in smiling movements must be measured when persons are known, by means other than the face, to be experiencing positive and negative affect (Ekman, Friesen, & Ancoli, 1980; Frank, Ekman, & Friesen, 1993).
4. Are there differences in heart rate that accompany nose wrinkling and upper lip raising versus opening the eyes and raising the brows? Facial behavior must be measured to identify the moments when these particular facial configurations occur in order to examine coincident heart rate activity (Levenson et al., 1990).

The preceding examples are not intended to convey the full range of issues that can be addressed only by measuring facial sign vehicles. They should, however, serve to illustrate the variety of questions that require this approach. One might expect the measurement of sign vehicles approach to have been followed often, as it is required for study of many different problems. However, there have been only a few such studies compared with the many that have measured the messages judged when viewing the face. It is much easier to perform the latter sort of study. The investigator need not tamper with the face itself, other than by picking some samples to show. Data are obtained quickly: One can measure observers' judgments much more quickly than one can describe reliably the flow and variety of facial movement.

Until the advent of FACS, an important obstacle to research measuring sign vehicles has been the lack of any accepted, standard, ready-for-use technique for measuring facial movement. Investigators who have measured facial movement have invented their techniques in large part de novo, rarely making use of the work of their predecessors. Some have seemed to be uninformed by the previous literature. Even the more scholarly have found it difficult to build on the methods previously reported, because descriptions of facial activity are often less clear than they appear on first reading. A facial action may seem to be described in sufficient detail and exactness until an attempt is made to apply that description to the flow of facial behavior. For instance, descriptions of brow motion that omit specific appearance changes in facial lines and furrows and in the appearance of the upper eyelid omit information that may be needed to discriminate among related but different facial actions. FACS addresses the need for a comprehensive system that can be readily learned, that is psychometrically sound, and that has high utility for various research applications.

Basis for Deriving Action Units

The anatomical basis of facial action (Figure 13.1) provides the basis for deriving units of behavior. With few exceptions, all people have the same facial muscles (Schmidt & Cohn, 2001). FACS action units are based on what the muscles allow the face to do. To determine the appearance changes associated with each muscle, Ekman and Friesen began by electrically stimulating individual muscles and by learning to control them voluntarily. The result is that each action unit is associated with one or more facial muscles.

In selecting facial actions, Ekman and Friesen (1978) used the criterion that observers were capable of reliably distinguishing all appearance changes resulting from the various muscles. If two appearance changes could not be reliably distinguished, they were combined, even if different muscles were involved. Conversely, some actions proved too subtle for reliable measurement. Visemes, for instance, are visually distinguishable phonemes (Massaro, Cohen, Beskow, Daniel, & Cole, 1998); with some exceptions, they are not included as AUs in FACS.

Facial Action Units

FACS 2002 specifies 9 action units in the upper face and 18 in the lower face. In addition, there are 14 head positions and movements, 9 eye positions and movements, 5 miscellaneous action units, 9 action descriptors, 9 gross behaviors, and 5 visibility codes. Action descriptors are movements for which the anatomical basis is unspecified. Upper and lower face AUs and head and eye positions are shown in Figure 13.2. Each one has both a numeric and verbal label and a specified anatomical basis in one or more facial muscles. With some exceptions, action units are organized by region of the face in which they occur. Brow action units, for instance, have AU labels 1, 2, and 4. There is no action unit 3 in FACS, although it is used to refer to a specific brow action in a specialized version of FACS intended for use with infants (Oster, 2001). Eye region action units have action unit labels 5, 6, and 7 and 41 through 46. For each action unit, FACS 2002 provides a detailed description, instructions on how to perform the action, and instructions for intensity scoring. For many action unit combinations, FACS 2002 covers these same topics and details subtle differences among related ones (e.g., AU 1+2 versus AU 1+2+4). Reference sections give information about AUs that might affect their scoring. By convention, when more than one action unit is present, they are listed in ascending order.

Appendix A describes some of the principal changes in action unit criteria and coding that occurred between FACS 1978 and FACS 2002. This material may be of particular interest to readers who have used the earlier version and are transitioning to FACS 2002.

FACS includes codes for head and eye positions. These action units often are omitted in FACS scoring. However, there is increasing evidence of their relevance to the interpretation of facial expression. Similar facial actions, such as smiling (AU 12), often vary in meaning depending on their temporal coordination with head motion. In embarrassment, for instance, smile intensity increases as the head pitches forward, and smile intensity decreases as the head pitches back toward frontal orientation (i.e., negative correlation; Cohn et al., 2004; Keltner & Buswell, 1997). FACS 2002 adds some specific codes for particular combinations of eye or head motion and facial action units, such as eyes (gaze) to the side occurring with AU 14, which may be a sign of contempt. Unless head and eye positions are scored, such relationships cannot be found.

Combinations of AUs that occur may be additive or nonadditive. In additive combinations the appearance of each action unit is independent, whereas in nonadditive combinations they modify each other's appearance. Nonadditive combinations are analogous to coarticulation effects in speech, in which one phoneme modifies the sounds of those with which it is contiguous. An example of an additive combination in FACS is AU 1+2, which often occurs in surprise (along with AU 5) and in the brow-flash greeting (Eibl-Eibesfeldt, 1989). The combination of these two action units raises the inner (AU 1) and outer (AU 2) corners of the eyebrows and causes horizontal wrinkles to appear across the forehead. The appearance changes associated with AU 1+2 are the product of their joint actions.

An example of a nonadditive combination is AU 1+4, which often occurs in sadness (Darwin, 1872/1998). When AU 1 occurs alone, the inner eyebrows are pulled upward. When AU 4 occurs alone, they are pulled together and downward. When they occur together, the downward action of AU 4 is modified. The result is that the inner eyebrows are raised and pulled together. This action typically gives an oblique shape to the brows and causes horizontal wrinkles to appear in the center of the forehead, as well as other changes in appearance.

Several chapters from FACS 2002 are viewable online at http://face-and-emotion.com/dataface/facs/manual/TOC .html.

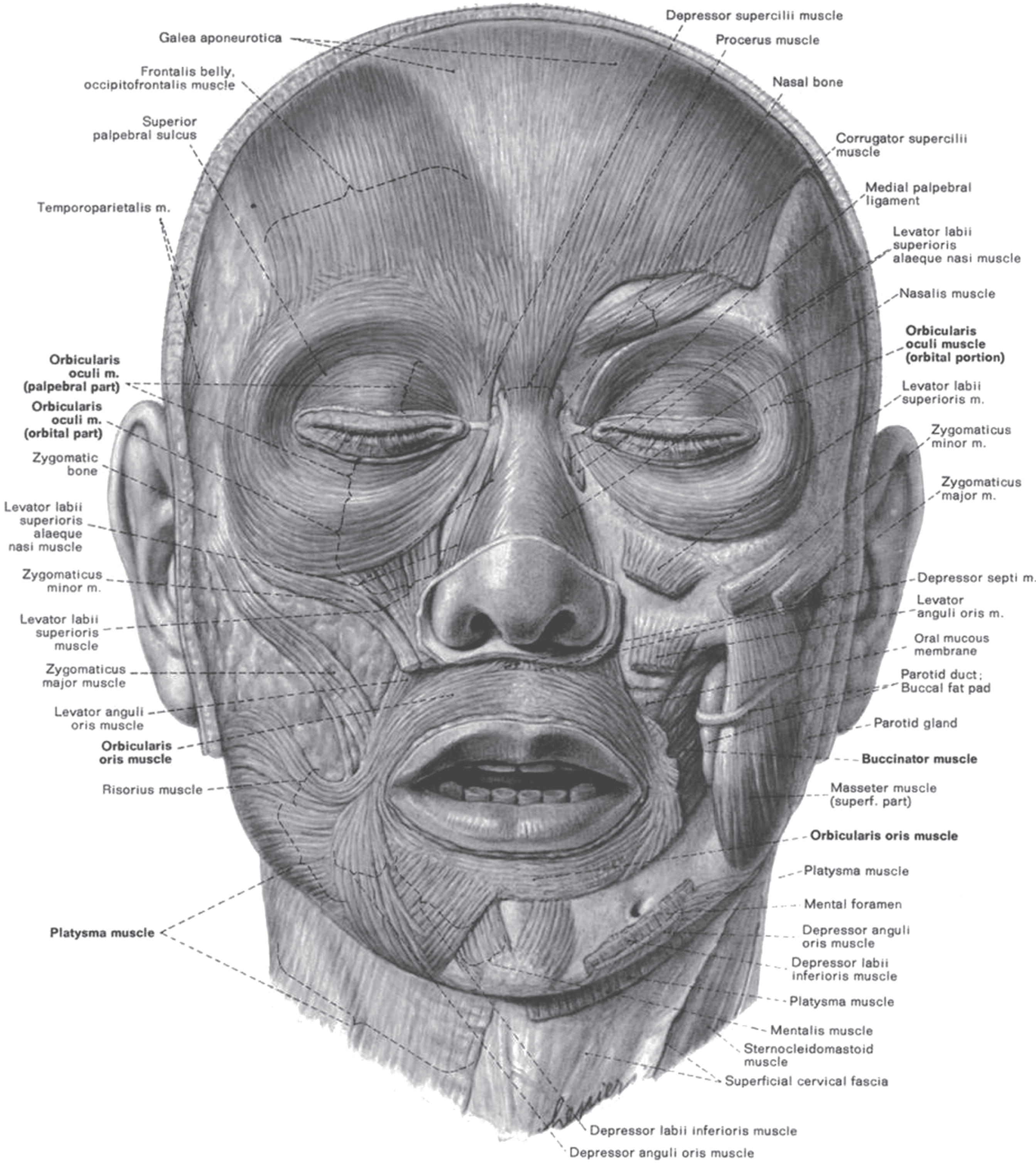


Figure 13.1. Muscles of the face (Clemente, 1997).

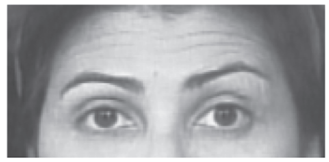
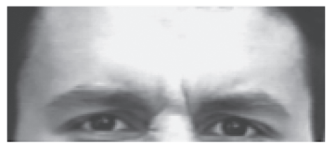
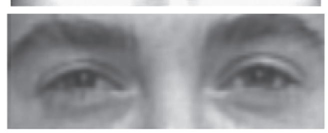
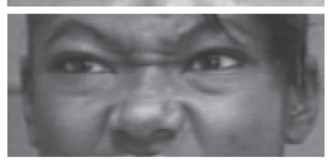
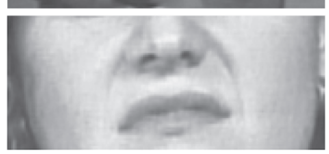
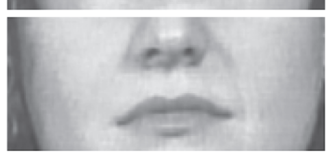
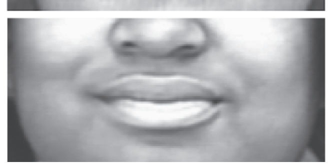
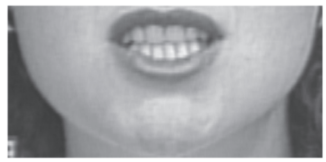
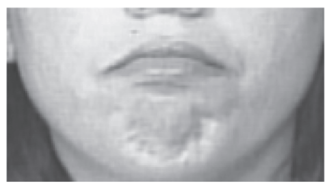
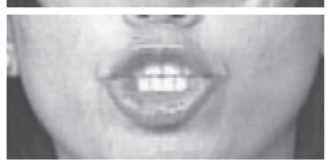
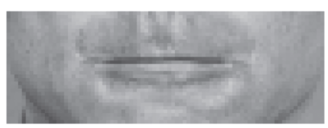
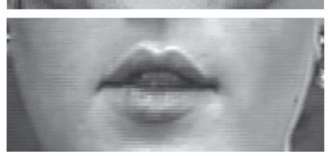
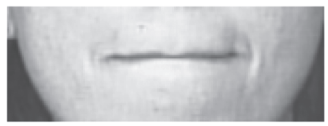
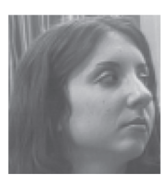
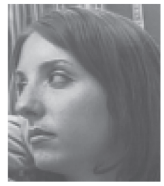
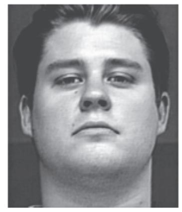
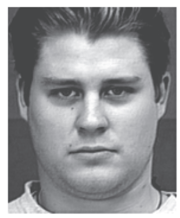
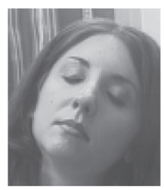
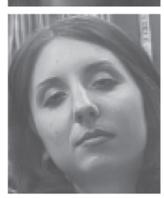
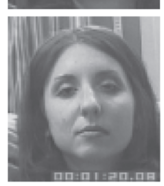
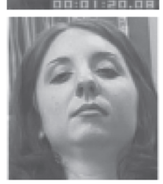
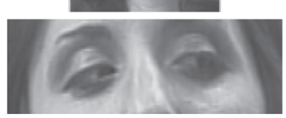
AU	Description	Facial muscle	Example image	Interrater agreement (Kappa coefficient) (tolerance window in seconds)			
				1/30th	1/6th	1/3rd	1/2
1	Inner Brow Raiser	*Frontalis, pars medialis*		.73	.79	.81	.83
2	Outer Brow Raiser	*Frontalis, pars lateralis*		.66	.71	.74	.76
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*		.58	.64	.67	.70
5	Upper Lid Raiser	*Levator palpebrae superioris*		.68	.76	.79	.82
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*		.72	.78	.82	.85
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*		.44	.49	.53	.56
9	Nose Wrinkler	*Levator labii superioris alaquae nasi*		.67	.76	.81	.83
10	Upper Lip Raiser	*Levator labii superioris*		.69	.76	.79	.81
11	Nasolabial Deepener	*Zygomaticus minor*		-	-	-	.97
12	Lip Corner Puller	*Zygomaticus major*		.67	.71	.74	.76

Figure 13.2. Action units of the Facial Action Coding System (Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 1978). *Note:* Interrater agreement is quantified using coefficient kappa, which controls for chance agreement. All kappa statistics are for spontaneous facial behavior and are from Sayette et al. (2001) except as noted in the text. Criteria for AUs 25, 26, 27, 41, 42, and 44 differ between FACS 1978 and FACS 2002. Please see appendix A for specifics. Images are from Kanade et al. (2000) with the exceptions of AU 23 and 24, which are from Ekman, Friesen, & Hager (2002).

AU	Description	Facial muscle	Example image	Interrater agreement (Kappa coefficient) (tolerance window in seconds)			
13	Cheek Puffer	*Levator anguli oris* (a.k.a. *Caninus)*		-	-	-	-
14	Dimpler	*Buccinator*		.59	.67	.72	.75
15	Lip Corner Depressor	*Depressor anguli oris (*a.k.a. *Triangularis)*		.54	.65	.69	.72
16	Lower Lip Depressor	*Depressor labii inferioris*		-	-	-	-
17	Chin Raiser	*Mentalis*		.55	.63	.66	.68
18	Lip Puckerer	*Incisivii labii superioris* and *Incisivii labii inferioris*		.65	.71	.74	.75
20	Lip stretcher	*Risorius* with *platysma*		.38	.47	.54	.60
22	Lip Funneler	*Orbicularis oris*		-	-	-	-
23	Lip Tightener	*Orbicularis oris*		.32	.41	.47	.53
24	Lip Pressor	*Orbicularis oris*		.50	.58	.62	.64
25	Lips parted	*Depressor labii inferioris* or relaxation of *Mentalis,* or		.57	.62	.65	.67

AU	Description	Facial muscle	Example image	Interrater agreement (Kappa coefficient) (tolerance window in seconds)			
		Orbicularis oris					
26	Jaw Drop	*Masseter,* relaxed *Temporalis* and *internal Pterygoid*		.65	.72	.76	.79
27	Mouth Stretch	*Pterygoids, Digastric*		-	-	-	.96
28	Lip Suck	*Orbicularis oris*		.61	.70	.76	.79
41	Lid droop	Relaxation of *Levator palpebrae superioris*		-	-	-	-
42	Slit	*Orbicularis oculi*		-	-	-	-
43	Eyes Closed	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*		-	-	-	-
44	Squint	*Orbicularis oculi, pars palpebralis*		-	-	-	.87
45	Blink	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*		-	-	-	.98
46	Wink	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*		-	-	-	-
51	Head turn left	---		-	-	-	-

AU	Description	Facial muscle	Example image	Interrater agreement (Kappa coefficient) (tolerance window in seconds)			
52	Head turn right	---		-	-	-	-
53	Head up	---		-	-	-	-
54	Head down	---		-	-	-	-
55	Head tilt left	---		-	-	-	-
56	Head tilt right	---		-	-	-	-
57	Head forward	---		-	-	-	-
58	Head back	---		-	-	-	-
61	Eyes turn left	---		-	-	-	-

This material conveys the thoroughness with which AUs are described, detailed information related to instruction, and subtle differences among AUs.

Scoring of Action Units

FACS provides researchers with a flexible range of options with respect to the level of detail with which action unit coding is performed. The options are not all mutually exclusive. Several may be combined, depending on the research question.

Comprehensive or Selective Coding

In comprehensive coding, each and every AU present in a chosen segment of facial behavior is coded. In selective coding, only predetermined AUs are coded; any others that appear are ignored. Each approach has its own advantages and disadvantages. The advantage of a comprehensive approach is that it allows researchers to analyze their data in more ways, to interpret null findings, and to make new discoveries. When a comprehensive approach is used, a null result can readily be interpreted as no differences between groups or conditions of interest. When a selective approach is used, the absence of differences between groups is open to question. There may be no difference between groups or conditions, or the investigator may have looked in the wrong places (i.e., chosen the "wrong" subset of actions to compare).

The primary drawback of comprehensive coding is that it is more labor intensive. A well-trained FACS coder can take about 100 minutes to code 1 minute of video data depending on the density and complexity of facial actions. The drawback of comprehensive coding is where the advantage of selective coding becomes apparent. Economy is the primary advantage of selective coding. Coding can be done more quickly because coders need to focus on only a subset of the facial actions.

Some research questions require a comprehensive approach to coding facial behavior. For example, an investigator who is interested in discovering whether there are unique facial behaviors that signal embarrassment would need to comprehensively code video segments of facial behavior during which participants reported feeling embarrassed and not embarrassed (Keltner, 1995). Keltner (1995) examined video records of people taken when they were reporting embarrassment, shame, or enjoyment. Comprehensive FACS coding of their facial expression enabled the discovery that self-reported embarrassment is uniquely associated with a particular sequence of action units: AU 12 followed by AU 24 and then by AU 51+54. Had the researcher selectively focused on only a subset of action units (for example, a smile with or without contraction of the muscle around the eye, AU 6, which is believed to distinguish between an enjoyment and a nonenjoyment smile), the discovery of the embarrassment display would not have been possible. Had a selective approach been used, they might have mistakenly concluded that there was no unique facial display of embarrassment. Exploratory studies benefit from a more comprehensive approach to coding.

Selective coding is best used when the investigator has strong hypotheses about specific action units or is interested only in specific facial regions (Messinger, Fogel, & Dickson, 1999; Prkachin, 1992). When selective coding is used, it is important to record which facial actions (AUs) were ignored. This record allows for more precise understanding of the results, so that readers/researchers can tell whether there was no result that involved the specific AUs or whether the AUs were not considered in the first place.

Presence/Absence or Intensity

Regardless of whether comprehensive or selective coding is used, researchers can determine the level of detail in their coding method. Most rudimentary is to code whether action units are present or absent. A different approach is taken when, in addition to coding presence or absence, coders also pay attention to the intensity or strength of the actions. FACS 2002 allows for five levels of intensity coding (A, B, C, D, and E), with A being the least intense (a trace) action and E the maximum strength of the action. Guidelines for intensity coding are somewhat subjective, however, and it may require special effort to establish and maintain acceptable levels of reliability, especially in the mid-range.

The importance of assessing intensity level depends on the nature of the study and the research question. In studies of pain, for instance, differences in intensity of facial actions have been found to vary among different types of pain elicitors (Prkachin, 1992). In related work, Deyo, Prkachin, and Mercer (2004) were interested in the functional relation between intensity of pain elicitors and intensity of selected action units; it was thus essential to code action unit intensity. As another example, Messinger (Messinger, 2002) proposed that mouth opening and cheek raising increase perceived intensity of infant's smile and distress independent of the strength of other actions (i.e., actual intensity of the AU 12 and AU 20, respectively).

Individual Action Units or Events

An event is a set of action units that overlap in time (i.e., co-occur) and appear to define a perceptually meaningful unit of facial action. AU 1+2+5 is an example of action units that frequently co-occur. Whether or not one AU begins or ends in advance of the others, they appear to constitute a single display. Rather than code each action unit independently, an investigator may wish to code or define them as an event. (A variant of this approach is found in Oster, 2001, and Oster et al., 1996). The guiding assumptions are:

1. Facial behavior occurs not continuously but rather as episodes (events) that typically manifest themselves as discrete events.
2. Action units that occur together are related in some way and form an event.

Event coding can be more efficient than coding single action units. Because action units are precoded into defined patterns, data reduction is facilitated as well. It also addresses the problem that some action units may linger and merge into the background. Events then can be coded independent of these longer lasting actions (e.g., a low-intensity AU 12 that may persist for some time). For investigators who wish to use event coding, FACS includes numerous suggestions for how to delineate events. These appear in FACS 1992, FACS 2002, and the Investigator's Guide that comes with FACS 2002 (Ekman et al., 2002).

There are some limitations to event coding. An event is a higher order perceptual unit than an action unit, and the rules or bases for identifying events are not as well defined. One basis is frequency of co-occurrence: AU combinations that are known to co-occur can be considered an event. A concern here is that normative data on co-occurrence rates are lacking and could be population specific, varying, for instance, with psychopathology or cultural differences. Another basis, suggested by the Investigator's Guide (Ekman et al., 2002), is a known association between specific AU combinations and specific emotions: An AU combination that is commonly associated with an emotion can be considered an event. This basis potentially violates the sign-based logic of FACS, which is to keep description separate from inference. To code some AUs as events based on their association with emotion while omitting others is more consistent with a judgment- than a sign-based approach. Studies that have utilized event coding typically do not report the basis on which they define an event. It may well be that association with emotion is not used often or even at all. But it is important that investigators realize the nature of event coding and recognize the potential bias it may introduce. The coder may impose organization where none exists or misidentify or omit events.

A related concern is that FACS guidelines for overlapping events may prove overwhelming, as there are quite a few exceptions to the rules and little or no rationale is provided. Sometimes an increase in AU intensity is treated as a separate event, whereas at other times it is treated as a "background" AU that is not coded in the subsequent event.

Unlike action unit coding, which is well validated, little is known about the psychometrics or perception of event coding. The literature in event perception is limited to the perception of complex scenes (Avrahami & Kareev, 1994; Newtson, Rindner, Miller, & Lacross, 1978; Zacks & Tversky, 2001). Studies remain to be done about how people segment facial behavior. Do people see facial behavior as episodic and event-like? Where and how do people segment the stream of facial behavior? Are people reliable in perceiving and determining the boundaries of these events? Carefully designed empirical studies are needed to answer such questions. In the meantime, given that delineating an event to a certain extent involves subjective judgments on the part of the coders, it raises a reliability issue of event segmentation, and hence it is important to establish and report coders' reliability on this issue. Ekman and Friesen found good reliability for event coding in developing FACS, but to our knowledge, no published studies of facial expressions that utilize the event-based approach have reported reliability for event determination, with the exception of emotion-specified aggregates, as presented subsequently.

Event coding has proven problematic when FACS coding is used for training computer algorithms in automatic FACS coding. Event coding typically ignores onset and offset times and only includes action units present at the "peak" of the event. Moreover, action units present at the peak may or may not be at their peak intensity at that moment, and event coding ignores other action units that may be present but not considered part of the event. Algorithms must learn not only to detect action units but also when to ignore ones that might also be present. The higher order, top-down decision making of event coding stands in sharp relief against the bottom-up sign-based coding of action units on which FACS is based.

Psychometric Evaluation

Reliability (Interobserver Agreement)

We report reliability, defined as interobserver agreement[1], for individual action units in spontaneous facial behavior when that has been established. Except as noted in the following, reliability for posed facial behavior is not reported here. In general, reliability for spontaneous facial behavior is less than that for posed behavior. In spontaneous facial behavior, camera orientation is less likely to be frontal, head motion larger and more varied, and face size smaller relative to the size of the image. In spontaneous facial behavior, action unit intensity may also be lower. All of these factors make reliability more difficult to achieve (Sayette, Cohn, Wertz, Perrott, & Parrott, 2001). Thus the reliability reported here represents a conservative estimate for investigators whose interest is in posed facial behavior.

Most studies that have reported reliability for FACS report average reliability across all AUs. As a consequence, there is no way to know which AUs have been reliably coded and which have not. Low reliability, especially for AUs occurring less frequently, may easily go unnoticed. This is of particular concern because these AUs often are of special interest (e.g., when occurring as microexpressions that qualify the interpretation of smiling; Ekman, Friesen, & O'Sullivan, 1988). When specific AUs are the focus of hypothesis testing, reliability at the level of individual AUs is needed. Otherwise, statistical power may be reduced by measurement error and, as a result, negative findings misinterpreted.

For individual AUs, at least four types of reliability (i.e., agreement between observers) are relevant to the interpretation of substantive findings. These are reliability for occurrence/nonoccurrence of individual AUs, as discussed earlier; temporal precision; intensity; and aggregates. Temporal precision refers to how closely coders agree on the timing of action units, such as when they begin or end. This level of reliability becomes important when hypothesis testing focuses on questions such as response latency. Action unit intensity becomes important when hypothesis testing focuses on questions such as whether intensity is related to subjective experience or individual differences. And, finally, in many studies, investigators are interested in testing hypotheses about emotion-specified expressions or events. The reliability of emotion–specified expressions or event coding will, of course, depend on the constituent AUs. By assessing the reliability of these aggregates directly, one can more accurately estimate their reliability.

The most systematic, large-scale investigation of FACS reliability in spontaneous facial behavior is that of Sayette and colleagues (Sayette et al., 2001). They evaluated each type of reliability across three studies of spontaneous facial behavior involving 102 participants, with approximately equal numbers of men and women. The studies induced change in facial expression by using one of three emotion inductions: olfactory stimulation, cue-induced craving for nicotine, and a speech task. Action units were comprehensively coded using FACS 1978 and FACS 1992. The number of frames (at 30 frames per second) for each action unit ranged from 800 (for AU 12) to less than 48 (e.g., AU 11). Action units occurring in less than 48 frames were not analyzed. Nineteen AUs met the authors' criterion of occurring in 48 or more video frames. We report reliability for these action units based on their findings.

To increase the number of AUs for which reliability is reported, we include findings from two other sources. One is from a study of spontaneous blinks (AU 45; Cohn, Xiao, Moriyama, Ambadar, & Kanade, 2003) in video data collected by Frank and Ekman (Frank & Ekman, 1997). In this study, male participants either lied or told the truth to an interviewer in a high-stakes deception interview. Some participants wore glasses, which made coding more challenging. The other source for reliability data is the Cohn-Kanade database (Kanade, Cohn, & Tian, 2000) of posed facial behavior (i.e., directed facial action tasks). We report reliability for AU 11, 27, and 44 from this database.

A caveat is that comparison coding in these two supplemental sources was not blind to the original codes. Instead, comparison coding was used to confirm or reject the original coding. Because comparison coding was not independent in coding AU 11, 27, 44, and 45, some caution must be used in interpreting the findings for these three AUs. We do not believe that bias was a significant factor, however, because independent comparison of three of these AUs (AU 27, 44, and 45) with automatic facial image analysis has been consistently high, especially for blinks (Cohn et al., 2003; Cohn et al., 1999; Tian, Kanade, & Cohn, 2002).

In all, the action units from the primary (Sayette et al., 2001) and supplementary (Cohn et al., 2003; Kanade et al., 2000) sources include those that have occurred most frequently in studies of both spontaneous and deliberate facial behavior. Reliability was assessed using coefficient kappa to control for agreement due to chance. With the exception of AUs 11, 27, and 44, all AUs occurred in spontaneous facial behavior.

Occurrence and Temporal Precision

Reliability for occurrence/nonoccurrence is reported in Figure 13.2. For the data from Sayette et al. (2001), reliability is reported for each of four tolerance windows. These tolerance windows represent the temporal precision with which action units were comprehensively coded, that is, from beginning to end. Only a single estimate is available for the four AUs from supplemental sources.

Using a one-half-second tolerance window, all but two action units (AU 7 and AU 23) had good to excellent reliability (see Figure 13.2). As the tolerance window decreased in size, reliability decreased; however, even at the smallest window, 11 of 19 AUs had excellent reliability. One of the AUs that had consistently low reliability was AU 23, which often is confused with AU 24; pooling them into a single category can improve reliability. Reliability for AU 7 was low. The revised coding of AU 7 in FACS 2002 should result in improved reliability for this AU, as it is now combined with AU 44.

Sayette et al. (2001) did not report reliability for specific phases of AUs, such as onset, offset, peak, intensity, or change in intensity; nor to our knowledge have such data been reported by other sources. The FACS 2002 Investigator's Guide presents data from a dissertation by Ancoli on temporal precision for AU onset and offset; these data show average agreement across all AUs for two samples. Temporal precision of onset and offset for individual AUs was not reported. Percent agreement in Sample 1 was low and may have reflected inexperience of the coders. In Sample 2, percent agreement (not kappa) within a one-tenth-second tolerance for onset and offset was 65% and 61%, respectively. Using a one-half-second tolerance window, the corresponding figures were 74% and 67%, respectively. More studies of this issue are needed.

Intensity

In Sayette et al. (2001), intensity was evaluated for AU 10, 12, 15, and 20. Reliability for intensity coding was not as high as what was found for occurrence/nonoccurrence and was better for AU 10 and AU 12 than for AU 15 and AU 20 (see Table 13.1). Although FACS 2002 provides for intensity coding of all action units, the current findings suggest that reliability for intensity coding may be problematic and that further work is needed.

Aggregates

Four aggregates, or emotion-specified events, were defined: positive and negative expressions, disgust, and sadness. Re-

Table 13.1
Kappa coefficients for 5-point intensity coding

	Tolerance Window (seconds)			
Action Unit	*1/30*	*1/6*	*1/3*	*1/2*
10	0.61	0.67	0.70	0.72
12	0.57	0.61	0.63	0.66
15	0.44	0.53	0.57	0.59
20	0.31	0.39	0.45	0.49

Adapted from Sayette et al. (2001).

liability for positive and negative aggregates and disgust was acceptable (kappa > 0.60) even at the shortest tolerance window (one-thirtieth of a second; Table 13.2). Reliability for sadness was acceptable only at one-third of a second or larger. The latter is likely to be an underestimate, however, in that the sadness aggregate occurred in only 37 frames.

Validity

Concurrent Validity

The validity of a technique designed to measure facial movement entails questions on a number of levels. Most specifically, validity requires evidence that the technique actually measures the behavior it claims to measure. When a technique claims to measure brow raise, are the brows actually raised, or is it just the inner corners that are raised? If the technique claims to measure the intensity of an action, such as whether the brow raise is slight, moderate, or extreme, do such measurements correspond to known differences in the intensity of such an action? The problem, of course, is how to know what facial action occurs and what criterion to utilize independently of the facial measurement technique itself. At least five approaches have been taken:

1. Performed action criterion. Ekman and Friesen trained people to be able to perform various actions on request. Records of such performances were scored without knowledge of the performances requested. FACS accurately distinguished the actions the performers had been instructed to make. These findings were replicated by Kanade and colleagues (2000).
2. Electrical activity criterion. Ekman and Friesen, in collaboration with Schwartz (Ekman, Schwartz, & Friesen, 1978) placed surface EMG leads on the faces of performers while the performers produced actions on request. Utilizing the extent of electrical activity observed from the EMG placements as the validity criterion, they found that FACS scoring of facial movement accurately distinguished the type and the intensity of the action.
3. Pixel-wise displacement criterion. Reed and Cohn (2003) compared maximum FACS intensity for AU 12 with automatic feature tracking of lip-corner displacement (see chapter 15, this volume, for a description of this measurement approach). Participants were young adults; spontaneous smiles occurred while they watched a video clip intended to elicit positive affect. FACS intensity for the onset phase of AU 12, measured on a 5–point scale, correlated .55 with pixel-wise lip-corner displacement, which suggests good concurrent validity.
4. Concurrent validity of FACS action units with automatic coding by computer-vision-based approaches. Concurrent validity has been demonstrated by several independent research groups. Examples from each of three separate groups include Bartlett et al. (1999), Cohn et al. (1999), and Pantic & Patras (2006).
5. Computer simulation. Work in computer graphics has shown that use of FACS action units in computer simulation can generate realistic movement of the target actions. FACS has provided an effective basis for facial animation, as well as for video encoding of facial images (Massaro et al., 1998; Parke & Waters, 1996).

Table 13.2
Kappa coefficients for emotion-specified combinations

		Tolerance Window (seconds)			
Action Unit Aggregates	*Frames*	*1/30*	*1/6*	*1/3*	*1/2*
Positive emotion	335	.71	.78	.81	.83
Negative emotion	313	.64	.74	.79	.82
Disgust	103	.75	.82	.85	.86
Sadness	37	.47	.61	.67	.73

Tabled values are from Sayette et al. (2001) and are based on 3-point intensity coding. Reliability of 5-point intensity scoring was not reported for emotion-specified expressions.

Stability

Several studies have found moderate stability in FACS action units and predictive validity for a wide range of personality and clinical outcomes. Cohn, Schmidt, Gross, and Ekman (2002) found moderate to strong stability in FACS action units over a 4-month interval; stability was sufficiently robust as to suggest that facial behavior could function as a biometric. Person recognition from FACS action units was comparable to that of a leading face recognition algorithm.

Utility

The utility of FACS has been shown in a wide range of studies with infants and adults in North America, Europe, and Asia. Research has included emotion and social processes, personality, psychopathology, pain, deception, perception, and biological bases. FACS has proven to be valid and useful for facial measurement in relation to each of these areas. Harker and Keltner (2001), for example, found that FACS action units predicted adjustment to bereavement, teacher ratings of problem behaviors, and marital adjustment over periods as long as 30 years. A review of this literature is beyond the scope of this chapter. A good introduction to this literature appears in Ekman and Rosenberg (2005).

Instruction

FACS 2002 (Ekman et al., 2002), a 370-page self-instructional text, is available on compact disk (CD). It comes with an Investigator's Guide that contains suggestions on the procedure for training FACS. For each action unit, there are detailed criteria described in the text and illustrated in photographic images and video examples. Frequently occurring action unit combinations, especially those involving nonadditive combinations as defined earlier, are included. Detailed tables are available that highlight similarities and differences between closely related action unit combinations. The material is thorough and well cross-referenced.

FACS 2002 may be learned through self-instruction or in groups with or without an expert leader. Training with a partner or in a group is preferable. Each person in a learning group can benefit from looking at others performing each AU, which is likely to contain idiosyncrasies in how an AU appears on individual faces. The group can also help to monitor reliability in practice scoring and is a motivational aid. Appendix B describes an intensive 1-week training course offered by Dr. Erika Rosenberg that has been proven highly successful. Here we present general guidelines for individual and group instruction.

In learning FACS, trainees are instructed to perform each AU and AU combination. This didactic, which Ekman and Friesen followed in developing FACS, helps the trainee master the mechanics of facial action and learn the idiosyncrasies in appearance changes associated with each AU. They are able to compare their own performances with those of other trainees and with the filmed and photographed demonstrations of each AU in the manual. Later on, this skill continues to be useful in coding, specifically to help coders determine whether the AUs are needed to visually duplicate the facial action being scored.

Consistent with the sign-based approach of FACS, the trainee is discouraged from thinking or talking about the meaning of AUs or AU combinations. With practice, coders are expected to think of facial action in terms of the AU labels (e.g., AU 1 and AU 2) and to forget about their meaning entirely.

The Investigator's Guide also provides detailed information about how a group of six coders progressed as they learned FACS. Trainees can use this information to compare their own performances and to guide them as they learn. For example, if a trainee makes more errors than those six coders did, then he or she may benefit from reviewing action units with which he or she experienced difficulty relative to the reference group.

Forty-six still photographs and 47 MPEG digital video clips are included for practice scoring. The answer key is provided in the Investigator's Guide, which also functions as a trainer's guide.

A computer program referred to as FACS Score Checker is provided with FACS 2002. This program serves to aid the trainee in comparing his or her scores for practice items with the criterion codes. It provides a more structured alternative to visual inspection of the correspondence between the trainee's scores and the criterion scores and contains features that help the trainee learn the scoring notation and avoid errors. It also provides a quantitative measure of success in learning to score with FACS.

FACS Score Checker can be used to save codes for the FACS Final Test, as described below. FACS Score Checker is written in Java 2 and can run on any computer on which the appropriate Java runtime is installed.

The time required to learn FACS is variable, depending on the number of hours per week that can be devoted to training, the availability of expert trainers, and individual differences among trainees. In our experience, approximately 3 months are required to become sufficiently proficient as to demonstrate mastery on the FACS Final Test. This would assume a training routine of weekly meetings and 3 to 4 days of practice per week.

FACS Final Test (Certification)

A certification test is available to demonstrate proficiency at the completion of training. Certification on this test is expected before one begins using FACS in research. The test consists of 34 short video segments in MPEG 2 format. The test items differ from the practice items provided in the FACS manual in a number of ways. The test video clips contain excerpts from actual conversations, and hence the expressions are spontaneous in nature. Head movements, speech, and nonfrontal orientation to the camera are common. These characteristics of the test items make them more difficult to code than the practice items in which speech is absent and camera orientation is frontal to the face. The test items represent the characteristics of actual facial behavior that might be typical in research materials. Following submission of the test scores, a trainee receives a reliability measure of his or her score compared with those of experts in FACS scoring. The trainee is also provided with a commentary on sources of errors in his or her scoring and suggestions for improvement. If his or her score is below criterion, retesting can be arranged. Procedures for obtaining and using the test are provided in the Investigator's Guide that accompanies FACS 2002.

Computer-Assisted FACS Coding

FACS comes with a score sheet that aids coders in organizing and recording their codes. This score sheet was designed

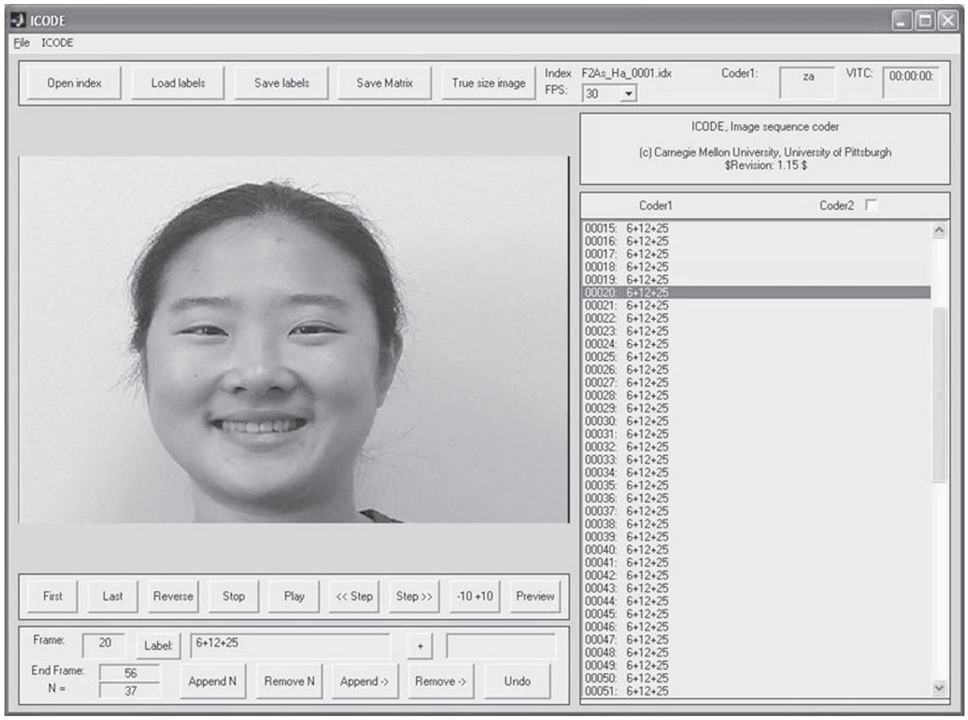


Figure 13.3. Snapshot of ICODE for computer assisted FACS coding. Available as shareware from http://www-2.cs.cmu.edu/~face/index2.htm

for coding single images that may occur alone or as part of a longer image sequence. Much early work in facial expression research focused on single images, for which the score sheet may have been sufficient. In practice, facial expression is dynamic and occurs in time. This dynamic nature of facial expressions is important for the recognition process (Ambadar et al., 2005), as well as for making finer distinctions among facial expression categories (Frank et al., 1993; Hess & Kleck, 1990). Timing of facial expressions influences judgments of genuineness among other dimensions (Krumhuber & Kappas, 2005; Schmidt, Ambadar, & Cohn, 2006). With the growing interest in analyzing the timing of facial expressions, a more sophisticated tool for coding dynamic facial expressions is needed. Using FACS to code video recorded facial expressions by hand (e.g., using paper and pencil or manual entry into a spreadsheet) is unnecessarily laborious and susceptible to error.

Computer-assisted coding enables precise control over video and eliminates the need to manually record time stamps or frame numbers, thus increasing efficiency and eliminating transcription errors. Image sequences may be viewed at variable speed forward and backward, and time stamps can be automatically logged whenever FACS codes are entered via the keyboard or other serial input device. Many of the systems that are available for computer-assisted coding also include a number of advanced data management, analysis, and display features. (See Bakeman, Deckner, and Quera (2005) for a more thorough discussion of computer-assisted coding.) Software for computer-assisted coding is available commercially and as shareware.

Examples of commercial systems are the Observer (Noldus, Trienes, Henriksen, Jansen, & Jansen, 2000; www.noldus.com), the Video Coding System (James Long Company; http://www.jameslong.net/), and INTERACT (Mangold Software and Consulting; http://www.mangold.de/english/intoverview.htm). Early versions of these and related systems are reviewed by Kahng and Iwata (1998). The most recent versions of Observer and Mangold INTERACT are well suited for complete FACS coding (all action units plus all modifiers).

Ordering Information for FACS 2002

FACS 2002 is available on CD-ROM from http://www.paulekman.com.

Appendix A: Changes in FACS 2002

FACS 2002 includes numerous changes, which are described here. The following list is not exhaustive but includes most changes.

General Changes

- FACS 2002 is available on CD and fully exploits this medium. The text is richly hyperlinked to related text, still images, and video. There are 135 still image examples and 98 video examples of single AU and AU combinations.

- General instructions for scoring intensity are presented in the introduction; criteria for specific action units are listed in corresponding sections of the manual. FACS uses conventions or rules to set thresholds for scoring action unit intensity. The criteria are set in terms of the scale of evidence of the presence of the AU. A distinction is made among a trace (A), slight (B), marked or pronounced (C), severe or extreme (D), and maximum (E) evidence.
- Change of the meaning of the notation "M." FACS 1978 listed the meaning of error messages from a Coder Check Program (CDRCHK), which came with the manual. All of these messages were numbered and had an "M" notation preceding the number. For example "M72" means "Two AU scores have been located that were not in numerical order, or a value greater than 74 has been identified. . . ." FACS 2002 does not list error messages. The notation "M" now is used to indicate "movement" of the head associated with other action units. For example "M59" indicates that the onset of 17 + 24 is immediately preceded, accompanied, or followed by an up-down head shake (nod).
- Changes in the key answer for the practice items. The Investigator's Guide of FACS 2002 includes a revision of the answer key for practice items. The revision reflects changes in scoring decisions, as well as changes in the rules (elimination of minimum requirements, reduction of co-occurrence rules, changes with AU 25, 26, 27 and 41, 42, etc).

Elimination of Action Units

- Elimination of AU 41 and 42. In FACS 1978, the numbers 41 and 42 were used to denote different actions of the same muscle in the eye region. Both AU 41 and AU 42 involve the relaxation of the muscle that, when contracted, raises the upper eyelid. With a relaxation of this muscle, the eyelid comes down, drooping (41) or almost entirely closed to just a slit (AU 42). In FACS 2002, these actions are no longer represented in different action units but are described as increasing degrees of upper eyelid lowering (AU 43b and 43d, respectively). However, the numbers 41 and 42 are used to indicate different strands of AU 4 (brow lowerer). Although only described in the Investigator's Guide, the recycling of these numbers to indicate completely different actions can create confusion and miscommunication. For this and other reasons, one should always report which version of FACS is used.
- Elimination of AU 44. In FACS 1978, AU 44 was used to denote a specific action of squinting. In FACS 2002, this action represents the maximum level of eyelid tightening as indicated by AU 7E. The number 44, however, is recycled and is used to indicate a different strand of AU 4, which is discussed in the Investigator's Guide (Ekman et al., 2002).

Elimination and Reduction of Scoring Rules

- Elimination of minimum requirements from all action units.
- Elimination of dominance rules.
- Reduction of Co-occurrence rule. Minimum requirements and co-occurrence rules were a pervasive aspect of the original FACS, designed to make scoring more deterministic and conservative, and thus more reliable. These expected benefits, however, were outweighed by difficulties in remembering and applying complicated rules, and experienced FACS coders proved able to make finer discriminations with greater reliability than these rules assumed. For these and other reasons, most co-occurrence rules and minimum requirements are eliminated in FACS 2002.

Modification in Scoring Rules

- FACS 2002 uses five levels of intensity coding for all AUs (except for head position AUs, in which the "A" level of intensity was treated as "Unscorable"). Intensity scoring was recommended only for some AUs in FACS 1978. However, in FACS 2002, guidelines for intensity scoring for all AUs are provided. Most intensity criteria refer to the degree of an appearance change or to the number of appearance changes. The intensity score for some AUs involves a criterion in terms of time duration or some other benchmark.
- A detailed description of AU combinations AU 6+12; 7+12; and 6+7+12 is added in FACS 2002. Previously, a strong action of AU 12 required that AU 6 be coded as well. In the 2002 version, this requirement is dropped. A careful inspection of the movement, the location of the crows' feet wrinkles, and the lowering of the outer corners of the brows are listed as necessary clues to determine whether AU 12 appears with or without AU 6 in any intensity level of AU 12. The possibility of AU combination 6+7+12 is for the first time introduced and described in detail in the current (2002) version of FACS. FACS 2002 notes, interestingly, that scoring a 6+7+12 is more likely than a 6+12.
- Revised scoring of AU 25, 26, and 27. FACS 2002 includes a more extensive description to separate the degree of mouth opening due to AU 25, 26, and 27. The previous version assumed that, by default, the action of jaw drop (AU 26) and mouth stretch (AU 27) included the separation of the lips (AU 25). In the case that the jaw drop was not accompanied by lip part, the score was indicated by S26 (Shut 26), which could be scored only when a movement of the jaw dropping

was visible. With this way of coding, if an action was coded as AU 26, it was assumed that the mouth was opened simultaneously. With AU 27 (mouth stretch), it was always assumed that the lips were also parted (AU 25). In FACS 2002, each action unit code is listed in the score whenever it is present. Therefore, a jaw drop that includes parting of the lips is coded AU 25 + 26. Likewise, mouth stretch action that includes lip part is to be coded as AU 25 + 27. Under the new rule, therefore, an action coded as 26 or 27 alone holds the assumption that the mouth remains closed. This change should eliminate much ambiguity. The manual also describes in detail the rules for scoring intensity with these mouth-opening AUs (25, 26, and 27).
- Timing information is used to differentiate some AUs, including 26 versus 27 and Unilateral 45 and 46.
- Revised rule for scoring AU 28. FACS 1978 recognized that the jaw is always lowered in AU 28, but scoring AU 26 was not recommended unless the lips were not touching. In FACS 2002, however, this rule is revised, and coders are to score AU 26+28 to indicate the jaw lowering action that allows the lip sucking action. To distinguish the lips touching from the lips parted, AU 25 should be added in the lip-part condition.
- In FACS 1992, the intensity scoring of AU 43 and 45 was determined by the intensity scoring of AU 6 and/or 7. In FACS 2002, there is no intensity scoring for AU 45 (blink), but for AU 43, the intensity-score guideline is provided; it is not determined by the intensity score of AU 6 or 7.

Addition of Action Unit or Behavior Category

- Two head movement codes are added. AU M59 (nod up and down accompanying AU 17 + 24) and AU M60 (shake side to side).
- Gross behavior codes are added. These are new codes that indicate behaviors of possible relevance to facial behavior. These new codes include: 40 (sniff), 50 (speech), 80 (swallow), 81 (chewing), 82 (shoulder shrug), 84 (head shake back and forth), 85 (head nod up and down), 91 (flash), and 92 (partial flash). No descriptions for these AUs are provided, however.

Appendix B: Description of Erika Rosenberg's FACS Workshop

Erika Rosenberg offers a 5-day intensive training workshop in FACS 2002.[2] The workshop takes participants through the entire manual and prepares them to take the final test for certification as a FACS coder. Traditionally, people have learned FACS via a minimum of 100 hours of self-instruction. The FACS workshop offers a dynamic group setting for learning FACS in about a week, with the benefit of guidance and feedback from an expert. FACS certification testing is done independently, after the workshop.

Basic Description of the Workshop

The 5-day workshop follows an intensive schedule of work and preparation. Participants read through the first three chapters of the FACS CD manual and do some preassigned practice scoring before coming to the workshop. This is necessary to get them thinking in terms of action units and to instill a sense of what the world of FACS is like. Part of the challenge of learning FACS is getting used to attending to feature details of the face (e.g., wrinkles, bulges, furrows, etc.) and understanding that FACS is a system of describing *movement.* She encourages students to think of themselves as "facial detectives" whose job it is to most efficiently describe observed changes in facial action.

The workshop schedule involves several activities: instruction in new actions, practicing AUs on one's own face, looking at other's faces, discussing difficulties in recognizing AUs, practicing coding with feedback, and evaluating progress daily. Rosenberg offers numerous clues to recognizing AUs—primarily by demonstrating on her own face and discussing subtleties of appearance changes included in the manual, but also by providing valuable field experience and tricks that are not available there. Part of the instruction is advice on how to use the manual most efficiently.

Daily quizzes and evening homework are an important component of the training, as they ensure that the students are staying with the class. The assignments also offer the instructor opportunity to evaluate each student's progress and problem areas. All homework and quizzes are graded.

By late in the week students have learned all AUs and head and eye positions (through chapter 9 in the manual), at which time they devote most of their time to practicing coding (with feedback), and discussion of common confusions and difficulties. Rosenberg provides real-life examples so that students can get a feel for what it is like to code spontaneous behavior in preparation for the FACS certification test. Also discussed are strategies for coding. Depending on the group, emotion interpretation may be included.

Effectiveness of the Workshop

Initial data suggested that people who train via the workshop pass the FACS final test the first time at the same rate as people who train via the traditional method. Everyone passes by the second time. Mark Frank—who has had more than 30 people trained in these workshops—reports that workshop trainees achieve intercoder agreement in his lab assessments more quickly and more often achieve reliability compared with people who have used the self-instruction method.

Contact Information

Further information may be obtained from http://www.erikarosenberg.com/.

Acknowledgments

Preparation of this manuscript was supported by NIMH grant MH 51435. The section on sign versus message judgment is from Cohn and Ekman (2005; used with permission of Oxford University Press). We thank Erika Rosenberg for her comments on an earlier version of this chapter.

Notes

1. We define *reliability* as agreement between two or more observers about the occurrence, intensity, and timing of action units. Alternatively, one might be interested in whether summary counts are consistent between coders. That is, are coders consistent in estimating the number of times an action unit occurs or its average intensity? Agreement is a more stringent measure in that coders must not only be consistent about the number of times an action unit occurs but must also agree on when each one occurred. Agreement may actually be quite low and yet reliability in the sense of consistency be quite high (Tinsley & Weiss, 1975). Agreement between coders is best quantified by coefficient kappa or a similar statistic that corrects for chance agreement. Reliability (consistency between coders) for summary counts is best quantified with intraclass correlation, which is mathematically related to kappa (Fleiss, 1981).

2. Appendix B is based on materials provided by Dr. Rosenberg and is used with her permission.

References

Ambadar, Z., Schooler, J., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions. *Psychological Science, 16,* 403–410.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. *Cognition, 53,* 239–261.

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

Bartlett, M. S., Ekman, P., Hager, J. C., & Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology, 36,* 253–263.

Breidt, M., Wallraven, C., Cunningham, D. W., & Buelthoff, H. H. (2003). Facial animation based on 3D scans and motion capture. *SIGGRAPH '03* Sketches & Applications, 0–0. (Eds.) Campbell, N. ACM Press, New York, NY, USA (07 2003).

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception.* New York: Oxford University Press.

Calder, A. J., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., & Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13,* 699–745.

Clemente, C. D. (1997). *Anatomy: A regional atlas of the human body* (4th ed.).Baltimore: Williams & Wilkins.

Cohn, J. F., & Ekman, P. Cohn, J. F. & Ekman, P. (2005). Measuring facial action. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of nonverbal behavior research* (pp. 9–64). New York: Oxford University Press.

Cohn, J. F., Reed, L. I., Moriyama, T., Xiao, J., Schmidt, K. L., & Ambadar, Z. (2004). Multimodal coordination of facial action, head rotation, and eye motion. Paper presented at the *IEEE International Conference on Automatic Face and Gesture Recognition,* 645–650.

Cohn, J. F., Schmidt, K. L., Gross, R., & Ekman, P. (2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. Paper presented at the *International Conference on Multimodal User Interfaces*, 491–496.

Cohn, J. F., & Tronick, E. Z. (1983). Three-month-old infants' reaction to simulated maternal depression. *Child Development, 54,* 185–193.

Cohn, J. F., Xiao, J., Moriyama, T., Ambadar, Z., & Kanade, T. (2003). Automatic recognition of eye blinking in spontaneous facial behavior. *Behavior Research Methods, Instruments, and Computers, 35,* 420–428.

Cohn, J. F., Zlochower, A. J., Lien, J. J. J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36,* 35–43.

Darwin, C. (1998). *The expression of the emotions in man and animals* (3rd ed.). New York: Oxford University Press. (Original work published 1872)

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., & Friesen, W. V. (1990). Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I. *Journal of Personality and Social Psychology, 58,* 330–341.

Deyo, K., Prkachin, K. M., & Mercer, S. R. (2004). Development of sensitivity to facial expressions of pain. *Pain, 107,* 16–21.

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

Ekman, P. (1964). Body position, facial expression and verbal behavior during interviews. *Journal of Abnormal and Social Psychology, 68,* 295–301.

Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology, 2,* 725–735.

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 169–222.). New York: Academic.

Ekman, P. (1980). Facial asymmetry. *Science, 209,* 833–834.

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

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384–392.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior. *Semiotica, 1,* 49–98.

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

Ekman, P., & Friesen, W. V. (1982). *Rationale and reliability for EMFACS*. Unpublished manuscript. University of California at San Francisco, Human Interaction Laboratory.

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39,* 1125–1134.

Ekman, P., Friesen, W. V., & Hager, J. C. (Eds.). (2002). *Facial Action Coding System* [E-book]. Salt Lake City, UT: Research Nexus.Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54,* 414–420.

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

Ekman, P., & Rosenberg, E. (Eds.). (2005). *What the face reveals* (2nd ed.). New York: Oxford University Press.

Ekman, P., Rosenberg, E., & Hager, J. (1998). *Facial Action Coding System Interpretive Database (FACSAID)*. Unpublished manuscript. University of California at San Francisco, Human Interaction Laboratory.

Ekman, P., Schwartz, G. E., & Friesen, W. V. (1978). *Electrical and visible signs of facial action.* San Francisco: University of California, Human Interaction Laboratory.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions.* New York: Wiley.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology, 72,* 1429–1439.

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology, 64,* 83–93.

Fridlund, A. J. (1992). The behavioral ecology and sociality of human faces. In M. S. Clark (Ed.), *Emotion* (pp. 90–121) London: Sage.

Harker, L., & Keltner, D. (2001). Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality and Social Psychology, 80,* 112–124.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1992). Primitive emotional contagion. In M. S. Clark (Ed.), *Emotion and social behavior* (Vol. 14, pp. 151–177). Newbury Park, CA: Sage.

Hess, U., & Kirouac, G. (2000). Emotion expression in groups. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotion* (pp. 368–381). New York: Guilford Press.

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology, 20,* 369–385.

International Organization for Standardization. (2002). *Overview of the MPEG-4 standard.* Retrieved from http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm

Izard, C. E. (1977). *Human emotions.* New York: Plenum.

Izard, C. E. (1979). *Facial expression scoring manual (FESM).* Newark: University of Delaware Press.

Izard, C. E. (1983). Maximally discriminative facial movement coding system (MAX). Unpublished manuscript, University of Delaware, Newark.

Izard, C. E., & Dougherty, L. M. (1981). Two complementary systems for measuring facial expressions in infants and children. In C. E. Izard (Ed.), *Measuring emotions in infants and children.* Cambridge,UK: Cambridge University Press.

Kahng, S., & Iwata, B. A. (1998). Computerized systems for collecting real-time observational data. *Journal of Applied Behavior Analysis, 31,* 253–261.

Kaiser, S. (2002). Facial expressions as indicators of "functional" and "dysfunctional" emotional processes. In M. Katsikitis (Ed.), *The human face: Measurement and meaning* (pp. 235–253). Dordrecht, Netherlands: Kluwer Academic.

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *IEEE Conference on Automatic Face and Gesture Recognition, 4,* 46–53.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122,* 250–270.

Keltner, D. (1995). Sign of appeasement: Evidence for the distinct display of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, 68(3), 441–454.

Kring, A. M., & Sloan, D. M. (1991). *The facial expression coding system: FACES.*Berkeley: University of California Press.

Krumhuber, E., & Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior, 29,* 3–24.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27,* 363–384.

Malatesta, C. Z., Culver, C., Tesman, J. R., & Shephard, B. (1989). The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development, 54.*

Massaro, D. W., Cohen, M. M., Beskow, J., Daniel, S., & Cole, R. A. (1998). Developing and evaluating conversational agents. *Workshop on Embodied Conversational Characters,* 137–148.

Matias, R., & Cohn, J. F. (1993). Are max-specified infant facial expressions during face-to-face interaction consistent with differential emotions theory? *Developmental Psychology, 29,* 524–531.

Messinger, D. S. (2002). Positive and negative: Infant facial expressions and emotions. *Current Directions in Psychological Science, 11,* 1–6.

Messinger, D. S., Fogel, A., & Dickson, K. L. (1999). What's in a smile? *Developmental Psychology, 35,* 701–708.

Newtson, D., Rindner, R., Miller, R., & Lacross, K. (1978). Effects of availability of feature changes on behavior segmentation. *Journal of Experimental Psychology, 14,* 379–386.

Noldus, L. P. J. J., Trienes, R. J. H., Henriksen, A. H. M., Jansen, H., & Jansen, R. G. (2000). The Observer Video-Pro : New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behavior Research Methods, Instruments, and Computers, 32,* 197–206.

Oster, H. (2001). *Baby FACS: Facial Action Coding System for infants and young children.*Unpublished manuscript, New York University.

Oster, H., Camras, L. A., Campos, J., Campos, R., Ujiee, T., Zhao-Lan, M., et al. (1996). The patterning of facial expressions in Chinese, Japanese, and American infants in fear- and anger- eliciting situations. Poster presented at the International Conference on Infant Studies, Providence, RI.

Oster, H., & Ekman, P. (1978). Facial behavior in child development. In A. Collins (Ed.), *Minnesota Symposium on Child Development* (Vol. 11). Hillsdale, NJ: Erlbaum.

Oster, H., Hegley, D., & Nagel, L. (1992). Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas. *Developmental Psychology, 28,* 1115–1131.

Pantic, M., and Patras, I. (2006). Dynamics of facial expressions: Recognition of facial actions and their temporal segments from profile image sequences. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 36 (2), 443–449.

Pantic, M., & Rothkrantz, M. (2000). Expert system for automatic analysis of facial expression. *Image and Vision Computing, 18,* 881–905.

Parke, F. I., & Waters, K. (1996). *Computer facial animation.* Wellesley, MA: Peters.

Prkachin, K. M. (1992). The consistency of facial expressions of pain. *Pain, 51,* 297–306.

Reed, I.L. & Cohn, J.F. (2003). Unpublished data. Affect Analysis Group, University of Pittsburgh, Pittsburgh, PA 15260.

Rinn, W. E. (1984). The neuropsychology of facial expression. *Psychological Bulletin, 95,* 52–77.

Rinn, W. E. (1991). Neuropsychology of facial expression. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior: Studies in emotion and social interaction* (pp. 3–30). New York: Cambridge University Press.

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.

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A psychometric evaluation of the Facial Action Coding System for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25,* 167–186.

Schmidt, K. L., Ambadar, Z., & Cohn, J. F. (2006). Timing of lip corner movement affects perceived genuineness of spontaneous smiles. Manuscript submitted for publication.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary perspectives in facial expression research. *Yearbook of Physical Anthropology, 116,* 8–24.

Sloan, D. M., Straussa, M. E., Quirka, S. W., & Sajatovic, M. (1997). Subjective and expressive emotional responses in depression. *Journal of Affective Disorders, 46,* 135–141.

Tao, H., Chen, H. H., Wu, W., & Huang, T. S. (1999). Compression of MPEG-4 facial animation parameters for transmission of talking heads. *IEEE Transactions on Circuits and Systems for Video Technology, 9,* 264–276.

Tian, Y., Kanade, T., & Cohn, J. F. (2005). Facial expression analysis. In S. Z. Li & A. K. Jain (Eds.), *Handbook of face recognition, 247*–276. New York: Springer.

Tian, Y., Kanade, T., & Cohn, J. F. (2002). Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, 5,* 229–234.

Tinsley, H., & Weiss, D. (1975). Interrater reliability and agreement of subjective judgements. *Journal of Counseling Psychology, 22,* 358–376.

Tomkins, S. S. (1962). *Affect, imagery, consciousness.* New York: Springer.

Tronick, E. Z., Als, H., & Brazelton, B. T. (1980). Monadic phases: A structural descriptive analysis of infant-mother face-to-face interaction. *Merrill-Palmer Quarterly of Behavior and Development, 26,* 3–24.

Van Swearingen, J. M., & Cohn, J. F. 2005). Depression, smiling and facial paralysis. In R. vanGelder (Ed.), *Facial palsies,* 373–386. Amsterdam, Amsterdam: Lemma Holland.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21.