

The Facial Expression Coding System (FACES):

A Users Guide

Ann M. Kring
Department of Psychology
3210 Tolman Hall
University of California, Berkeley
Berkeley, CA 94720

Denise Sloan
Department of Psychology
Weiss Hall
Temple University
Philadelphia, PA 19122

****Unpublished manuscript - do not cite without permission of the first author**

The Facial Expression Coding System (FACES): A User's Guide

Facial expression of emotion is of great interest to many researchers. It has been studied in connection with subjective emotional experience, physiological arousal, and communication to name but a few areas. Interest in facial expression has a rich history dating back to the mid 19th century (Piderit, 1858, 1888; Gratiolet, 1865). Perhaps the most influential of these early theorists was Charles Darwin. In his book, *The Expression of the Emotions in Man and Animals*, Darwin (1896) argued that facial expressions were universal and innate characteristics. In the infancy of the science of psychology, William James hypothesized that facial expression played a causative role in the experience of emotion. In fact, according to James, changes in the facial musculature comprised a large portion of emotional state. James' ideas about emotion set forth a tradition of scholarly debate about the role of facial expression in emotion that continues today. In the early 1960's, Tomkins (1962, 1963) proposed what has become known as the facial feedback hypothesis. Stated succinctly, facial feedback theory holds that facial expression provides feedback which in turn produces the emotion. A tradition of research investigating the facial feedback hypothesis (see Adelman & Zajonc, 1989 for a review) has ensued, but several unanswered questions remain regarding the mechanisms and functions of facial expressions of emotion.

Systems for Coding Facial Expression

Ekman and Friesen (1976, 1978) were pioneers in the development of measurement systems for facial expression. Their system, known as the Facial Action Coding System or FACS, was developed based on a discrete emotions theoretical perspective and is designed to measure specific facial muscle movements. A second system, EMFACS, is an abbreviated version of FACS that assesses only those muscle movements believed to be associated with emotional expressions. In developing these systems, Ekman importantly distinguishes between two different types of judgments: those made about behavior (measuring sign vehicles) and those that make inferences about behavior (message judgments). Ekman has argued that measuring specific facial muscle movements (referred to as action units in FACS) is a descriptive analysis of behavior, whereas measuring facial expressions such as anger or happiness is an inferential process whereby assumptions about underlying psychological states are made. It is important to point out, as Ekman does, that any observational system requires inferences about that which is being measured. Other available systems have been designed to measure either specific aspects of facial behavior (e.g., Ermiane & Gergerian, 1978; Izard, 1979; see Ekman (1982) for a selective review) or more generally defined facial expressions (e.g., Notarious & Levenson, 1979).

Why a New System?

The primary reason for developing a new system was based on the perceived need for a facial coding that is theoretically aligned with a dimensional model of emotion. Several researchers have argued that affective expression consists of two broad dimensions: valence and arousal (e.g., Russell, 1980; Schlosberg, 1952). Similarly, researchers have argued that emotional experience variance is also best captured by two dimensions (e.g., Larsen & Diener,

1992; Watson, Clark, & Tellegen, 1988). Most currently available coding systems of facial behavior are based on discrete emotion theory and are designed to measure a number of specific or basic emotions. Although some might argue that these systems can be considered "dimensional" to the extent that discrete categories can be combined to form dimensions, this approach is inconsistent with the empirical literature upon which dimensional models of emotion have developed. The Facial Expression Coding System (FACES) was designed as a dimensional measure of facial behavior.

Second, while the Ekman and Friesen systems have been the pacesetters for studying facial expression of emotion, they are not without cost. It takes a great deal of time to train coders to use the system accurately and reliably (Ekman (1982) estimated approximately 100 hours were needed for training). Additionally, coding time can be quite extensive and as a result, often only small segments of participants' facial behavior are coded with FACS. EMFACS is somewhat more economical in that coders are not required to detect each muscle change but rather decide if a group of changes presumed to be associated with particular emotions have occurred. Being restricted to examine small portions of a participants data, although useful if researchers are interested in identifying specific responses to specific stimuli, can also be problematic. First, examining small segments may obscure an examination of the natural unfolding of expressive behavior over time. Second, selection of these segments most often requires *a priori* decisions about which segment is likely to produce the most expressive behavior. Selecting segments which maximize the likelihood of expressive behavior for all participants can be quite difficult.

An Overview of FACES

The Facial Expression Coding System (FACES) was developed as a less time consuming alternative to measuring facial expression that is aligned with dimensional models of emotion. The system provides information about the frequency, intensity, valence, and duration of facial expressions. The selection of the variables included in the system was based on theory and previous empirical studies. Adopting the descriptive style of Ekman and similar to the work of Notarius and Levenson (1979), an expression is defined as any change in the face from a neutral display (i.e., no expression) to a non-neutral display and back to a neutral display. When this activity occurs, a frequency count of expressions is initiated. Next, coders rate the valence (positive or negative) and the intensity of each expression detected. Notice that this is quite different from assigning an emotion term to each expression. While FACES requires coders to decide whether an expression is positive or negative, it does not require the application of specific labels. There is support in the literature for this approach, often referred to as the cultural informants approach (Gottman & Levenson, 1985). That is, judgments about emotion, in this case whether an expression is positive or negative, are made by persons who are considered to be familiar with emotion in a particular culture. In addition to valence and intensity, coders also record the duration of the expression. Finally, a global expressiveness rating for each segment is made, and judgements about specific emotions expressed throughout the segment can also be obtained.

How to Use FACES

FACES was initially developed to measure facial expressions in response to five minute film clips. The system can be adapted to other applications, however, and attempts to represent the broad applicability of the system are made throughout the manual. Generally speaking, the system allows for the examination of a participant's entire record of expressive behavior. When we videotaped participants viewing emotional films, the soundtrack from the movie was not included on their videotapes. Thus, coders only viewed participants facial reactions to the films. We have typically had two raters coding each participant. As will be discussed below, reliability for FACES has routinely been very high.

Detecting an Expression

While viewing a participant's record, an expression is detected if the face changes from a neutral display to a non-neutral display and then back to a neutral display. It is important to note, however, that a facial display may not always return to the original neutral display but may instead return to a display that, although neutral, does not exactly resemble the prior neutral expression. Additionally, if after a participant displays a shift from a neutral to non-neutral display and, instead of returning to a neutral display, shows a clear change in affective expression, this change is counted as an additional expression. For example, if while smiling, a participant then raises his or her eyebrows and stops smiling, indicating more of a surprised look, two expressions will be coded.

Duration

Once an expression has been detected, the duration is measured. For convenience, a time-mark in seconds should be included on participants' videotape. The duration measurement should start as soon as the participant changes from a neutral to non-neutral display. This time should be recorded on a coding form (sample coding forms are presented in the Appendix). The duration measurement should stop as soon as the participant changes back from a non-neutral to neutral display, and the time should be recorded on the coding form. The duration in seconds can then be calculated by subtracting the beginning time from the end time and then recorded on the coding form. Mean duration can be calculated by dividing the total duration by the frequency of expressions. Typically this is done separately for positive and negative expressions.

Valence

Next, the coder must decide whether the expression was positive or negative and make the appropriate notation on the coding form. If there is doubt as to whether the expression is positive or negative, a comprehensive list of affect descriptors is presented in the Appendix. Extensive research (Russell, 1980; Watson & Tellegen, 1985) has established these descriptors as either positive or negative. They are provided simply as a guide for coders in determining the valence of an expression. Coders are not asked to supply a descriptor for each expression detected.

Intensity

Intensity ratings for an individual expression range from one to four (1=low, 2=medium, 3=high, 4=very high). The low rating is given for those expressions that are mild, such as a smile where a participant slightly raises the corners of his/her mouth but does not show the teeth, and very little movement around the eyes occurs. The medium rating is given for those expressions where a participant's expression is more moderate than mild in intensity, such as a smile bordering on a laugh, with the eyebrows slightly raised and the lips apart, exposing the teeth. The high rating is given for an expression that involves most, if not all, of the face, such as laughing with an open mouth and raising the eyebrows and cheeks. The very high rating is reserved for those expressions that are very intense. An example of such an expression is one where a participant is undeniably laughing, with the mouth completely open with the eyebrows and cheeks substantially raised.

Summarizing the data

When film clips are the stimuli, we have found it useful to provide summary information at the end of each film clip. Specifically, two, subjective global ratings are taken: judgments about the specific emotion(s) being expressed and a judgment about the overall level of expressiveness. Additionally, summary information is calculated for the frequency, intensity, and duration measures for both positive and negative expressions. Two sample summary sheets are found in the Appendix.

Predominant Emotion Expressed

Although not a primary focus of the system, we have used two different rating schemes to assess more specific judgements of individual emotions: a forced choice rating and Likert format ratings. Using the forced choice method to determine the predominant emotion expressed, the coder should look over the coded expressions for the entire segment to obtain an assessment of whether the participant was expressing predominantly positive or negative emotions. Then, the coder is required to choose one of seven emotions on the summary form (happiness, sadness, disgust/fear, interest, neutral/indifferent, surprise, or anger). These were chosen as manipulation checks for the emotional film clips and can certainly be modified for different applications. This can be a difficult item to code. For example, a participant who was expressive during the segment can still obtain a global rating of neutral/indifferent if the expressions were all low in intensity and short in duration.

Using Likert scales, coders are required to rate, using a six point scale (1 = not at all to 6 = very much), the degree to which each specific emotion (e.g., happiness, sadness, amusement, fear, disgust, anger, interest) was expressed during the segment.

Level of Expressiveness

This rating is the coder's global assessment of expressiveness during a segment. Before making this assessment, the coder should look at all the individual expression ratings during

the segment. That is, this rating requires consideration of individual ratings of valence, duration, and intensity. The global rating of expressiveness ranges from one to five (1=low, 2=fairly low, 3=moderate, 4=fairly high, 5=high). A low rating would be given to a participant who had none or few expressions all of which were short and low in intensity. In contrast, a high rating would be assigned to a participant who had many highly intense and longer expressions.

Summary Measures

The total number of expressions is computed by simply counting the frequency of positive and negative expressions and recording these on the summary form. Similarly, the duration for expressions is computed by adding together the seconds for the positive and negative expressions (computed separately) and recording them on the summary form. Calculating mean duration is accomplished by dividing the total duration of positive expressions by the number of positive expressions. Mean negative duration is calculated by dividing the total negative duration by the number of negative expressions. Calculating the mean positive intensity requires that the positive intensity ratings be added together and divided by the number of positive expressions. In the same fashion, the mean negative intensity is calculated by dividing the sum of the negative intensity ratings by the number of negative expressions. The means for duration and intensity can be included on the summary sheet, or can be easily calculated using whatever statistical package you use (e.g., SPSS, BMDP, SAS, SYSTAT).

Things to Watch out for:

Experience with the coding system tells us that there are a number of things that can be problematic for coders if they are not discussed ahead of time. Below, we provide a list of the most common problems. This list is necessarily tied to our application and thus may not be applicable in other studies or settings. These suggestions are offered as guides, not absolute solutions, for coders.

Shifting Body Positions

A coder may sometimes mistake a change of body position for a change from a neutral to non-neutral facial display. Coders must take special care to ensure that the face changes in addition to the body posture shift in order to record that an expression has occurred.

Not paying attention

The coder should not code any expressions if the participant does not appear to be paying attention to the stimuli. Although this can be difficult to determine, if the participant is looking down or away from the stimulus that is being presented, it is likely that he or she is not attending. We have also employed separate ratings of attention in order to assess this more systematically. Depending on the application and participant population, this may be advisable.

Hand covering part of face

If a participant's hand is covering part of the face, the coder may unfortunately need to rely on the other parts of the face to detect the occurrence of facial expressions. For example, if the participant is covering the mouth area, the coder will need to pay special attention to the eye, nose, forehead, and cheek areas to code expressions.

Eye glasses

If the participant is wearing eye glasses, the coder may find it difficult to examine the participant's eyes during facial expressions. In this situation, the coder is encouraged to examine as best as possible eye movements (e.g., eyebrows raised above the eyeglass frame) as well as other areas of the face when determining whether or not an expression has occurred.

Contact lenses

If the participant is a contact lens wearer, chances are good that he or she may have eye movements related to the lenses and not to facial expressions per se. If possible, determine ahead of time if the participant wears and/or experiences any problems with contact lenses. If repetitive movements (e.g., blinking) occur that do not appear to be tied to the stimulus presentation, these should not be coded as facial expressions. Determination of this can be difficult and is best established by observing several occurrences of such movements across stimulus presentations.

Gum chewing

Participants chewing gum can present a sticky problem for coders. Gum chewing may actually inhibit natural expressive displays. The best solution here is to make sure a participant removes gum prior to the beginning of stimulus presentation.

Talking

Talking during a study can be problematic if more than one participant is being run through a study or if an experimenter is in the room. The best advice is to strongly encourage participants to refrain from talking during the experiment. In the event that coders are faced with rating a segment in which a participant is talking, attempts should be made to identify an expression independent of the talking. For example, if a participant smiles and then begins talking, the smile should be recorded as an expression. If on the other hand, the participant begins talking and has clearly diverted attention from the stimulus presentation, and smiles, it should not be recorded as an expression.

Facial tics

Occasionally, a participant may repeatedly display facial movements that do not appear to be expressions of emotion and are instead facial tics. As cited above, contact lens wearers

may have eye movements that are related only to the lenses. Other people may have other repetitive facial movements. These may not be obvious initially, but after viewing several minutes of a participant's record, they may become more prominent. A special case involves psychiatric patients with tardive dyskinesia. Patients who have taken neuroleptic medication for long periods of time may develop this very unfortunate side effect. Tardive dyskinesia involves uncontrollable repetitive movements that may involve facial muscles, most often those around the mouth. Any work done with psychiatric patients should involve careful assessment of these symptoms.

Assessing Rater Agreement

Since its inception, reliability for coders using FACES has been calculated using the intraclass correlation coefficient. The intraclass correlation coefficient is the correlation between one measurement (e.g., ratings of facial expressions) on a target and another measurement made on that same target (Shrout & Fleiss, 1979). More specifically, following the Case 2 study described by Shrout & Fleiss (1979), the coders (judges) are considered to be selected from a random sample of judges, and each judge rates each subject or target. That is, it is assumed that FACES can be used effectively by any set of coders. The formula used to calculate the ICC is derived from the components of a two-way ANOVA (Subjects x Coders) which partitions the within-target sum of squares into a between-coders sum of squares and a residual sum of squares. Because the variance due to coders is not ignored, the coefficient can be interpreted as an index of agreement rather than consistency (Shrout & Fleiss, 1979). As such, the formula is:

$$ICC = \frac{BMS - EMS}{BMS + (k - 1)EMS + k(CMS - EMS)/n}$$

where:

BMS = between subjects mean square

EMS = residual mean square

CMS = between coders mean square

k = number of coders

n = number of subjects

In our applications, using trained undergraduate and graduate students as coders and with varied participant populations (e.g., undergraduates, adult community residents, psychiatric patients), the agreement has been very high, ranging from .70 to .99.

References

- Adelmann, P. K., & Zajonc, R. B. (1989). Facial efference and the experience of emotion. *Annual Review of Psychology, 40*, 249-280.
- Darwin, C. R. (1896). *The expression of emotions in man and animals*. New York: Appleton.
- Ekman, P. (1982). Methods for measuring facial action. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research*. Cambridge: Cambridge University Press.
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Journal of Environmental Psychology, 1*, 56-75.
- Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System*. Palo Alto, CA: Consulting Psychological Press.
- Ermiane, R. & Gergerian, E. (1978). *Atlas of Facial Expressions. Album des expressions du visage*. Paris: La Pensee Universelle.
- Gottman, J. M., & Levenson, R. W. (1985). Assessing the role of emotion in marriage. *Behavioral Assessment, 8*, 31-48.
- Gratiolet, P. (1865). *De la physionomie et des mouvements d'expression*. Paris: Hetzel.
- Izard, C. (1979). *The maximally discriminative facial movement coding system. (MAX)*. Unpublished manuscript. Available from Instructional Resources Center, University of Delaware, Newark.
- Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In Margaret Clark (Ed.), *Review of Personality and Social Psychology, 13*, 25-59.
- Notarius, C., & Levenson, R. (1979). Expressive tendencies and physiological response to stress. *Journal of Personality and Social Psychology, 37*, 1204-1210.
- Piderit, T. (1858). *La Mimique et al physiognomie*. Paris: Alcan.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178.
- Schlosberg, H. (1952). The description of facial expression in terms of two dimensions. *Journal of Experimental Psychology, 44*, 229-237.

- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Tomkins, S. S. (1962). *Affect, Imagery, Consciousness: Vol 1. The Positive Affects*. New York: Springer.
- Tomkins, S. S. (1963). *Affect, Imagery, Consciousness: Vol 2. The Negative Affects*. New York: Springer.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219-235.

Appendix

FACES Coding Sheet

Participant _____	Film # _____	Rater _____	Page _____
<hr/>			
Time start _____	Time end _____	Duration _____	
Valence:	Positive ____	Negative ____	
Intensity:	low 1	medium 2	high 3
			very high 4
<hr/>			
Time start _____	Time end _____	Duration _____	
Valence:	Positive ____	Negative ____	
Intensity:	low 1	medium 2	high 3
			very high 4
<hr/>			
Time start _____	Time end _____	Duration _____	
Valence:	Positive ____	Negative ____	
Intensity:	low 1	medium 2	high 3
			very high 4
<hr/>			
Time start _____	Time end _____	Duration _____	
Valence:	Positive ____	Negative ____	
Intensity:	low 1	medium 2	high 3
			very high 4
<hr/>			
Time start _____	Time end _____	Duration _____	
Valence:	Positive ____	Negative ____	
Intensity:	low 1	medium 2	high 3
			very high 4
<hr/>			

Positive and Negative Affect Descriptors

Positive

Happy
Delighted
Glad
Amused
Pleased
Content
Satisfied
Calm
Serene
Excited
Astonished
Cheerful
Surprised
Active
Content

Negative

Miserable
Distressed
Annoyed
Jittery
Nervous
Angry
Gloomy
Anxious
Afraid
Tense
Alarmed
Frustrated
Disgusted
Depressed
Hostile

LIKERT FORMAT SUMMARY SHEET:

FACES Summary Sheet

Participant _____ Film # _____ Rater _____

Please rate the degree to which the participant expressed each of the following emotions using the scale below:

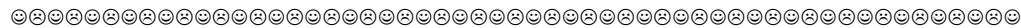
Not at all = 1 slightly = 2 somewhat = 3 moderately = 4 quite a bit = 5 very much = 6

Interest ___ sadness ___ happiness ___ anger ___ fear ___ amusement ___ disgust ___

What is the overall level of expressiveness for this person for this film?

low fairly low medium fairly high high
1 2 3 4 5

Number of positive expressions _____ Number of negative expressions _____
Mean positive intensity _____ Mean negative intensity _____
Duration of positive expressions _____ Duration of negative expressions _____



Participant _____ Film # _____ Rater _____

Please rate the degree to which the participant expressed each of the following emotions using the scale below:

Not at all = 1 slightly = 2 somewhat = 3 moderately = 4 quite a bit = 5 very much = 6

Interest ___ sadness ___ happiness ___ anger ___ fear ___ amusement ___ disgust ___

What is the overall level of expressiveness for this person for this film?

low fairly low medium fairly high high
1 2 3 4 5

Number of positive expressions _____ Number of negative expressions _____
Mean positive intensity _____ Mean negative intensity _____
Duration of positive expressions _____ Duration of negative expressions _____

FORCED CHOICE FORMAT SUMMARY SHEET

SUMMARY SHEET

Participant _____

Rater _____

Clip # _____

What was the predominant emotion being expressed throughout the clip?

- A. Interest
- B. Sadness
- C. Anger
- D. Surprise
- E. Fear/Disgust
- F. Neutral/Indifferent
- G. Happiness

What is the overall level of expressiveness for this person for this clip?

low	fairly low	medium	fairly high	high
1	2	3	4	5

Number of positive expressions _____

Number of negative expressions _____

Mean intensity--positive _____

Mean intensity--negative _____

Duration of positive expressions _____ (in seconds)

Duration of negative expressions _____ (in seconds)