In Scherer, K.R. and Ekman, P. (Eds.),
HANDBOOK OF METHODS IN NONVERBAL
BEHAVIOR RESEARCH. New York: Cambridge
University Press, 1982. Pp. 45- 135.

2. Methods for measuring facial action

PAUL EKMAN

2.1. Introduction

Of all the nonverbal behaviors – body movements, posture, gaze, proxemics, voice – the face is probably the most commanding and complicated, and perhaps the most confusing. In part, the face is commanding because it is always visible, always providing some information. There is no facial equivalent to the concealment maneuver of putting one's hands in one's pockets. Whereas sounds and the body movements that illustrate speech are intermittent, the face even in repose may provide information about some emotion or mood state. Many nonverbal behaviors simply do not occur when a person is alone, or at least do so very rarely. For example, it would be unusual for someone to shrug or gesture hello when totally alone. Yet facial expressions of emotion may be quite intense even when a person is alone. They are not occasioned only by the presence of others. In fact, social situations can dampen facial expression of emotion (Ekman, 1972; Ekman & Friesen, 1975, chap. 11).

The face is commanding also because it is the location for the senses of smell, taste, sight, and hearing. It is the site of the intake organs for inputs of air, water, and food necessary to life. It is the output source for speech, and what we hear in part is determined by the lip movements we see with the speech (McGurk & MacDonald, 1976). It commands attention because it is the symbol of the self. The faces of those we care about are hung on walls, displayed on desks, carried in wallets.

Multimessage–multisignal system

This commanding focus of attention is quite complex. The face can be considered as a multimessage, multisignal semiotic system (Ekman, 1978). The face conveys not only the message of individual identity, but also messages about gender and race. Certain changes in the face reveal,

more or less truthfully, age. There are standards for beautiful and ugly, smart and stupid, strong and weak faces. And apart from stereotypes, there have been claims for accurate information about personality traits, psychopathology, and ir telligence from facial behavior.

These different messages (identity, gender, beauty, traits, etc.) have as their source one of four types of facial signal systems: static, slow, artificial, and rapid. *Static* signs include the size, shape, and relative locations of the features and the contours produced by the underlying bony structure. These static signs are the likely vehicles for transmitting information about identity and beauty. Examples of *slow* sign vehicles would be the accumulation of wrinkles, pouches, and bags, which occur with and convey information about age. *Artificial* signs, such as cosmetics and plastic surgery, attempt to disguise these slow age signs. The *rapid* signs include the actions produced by the muscles (typically called expressions), as well as changes in muscle tonus, blood flow, skin temperature, and coloring.

Most research on the face has focused just upon these rapid signs, in particular, the momentary movements of the face and the muscle tonus changes as sign vehicles for information about emotion and mood. Rapid signs may also be relevant sources for other messages, for correct or incorrect information about traits, attitudes, personality, and so on.[1] Our focus in this chapter is upon the methods for measuring momentary facial movement (expressions). Later such methods will be compared with electromyographic measures of muscular activity.

Two methodological approaches

Ekman and Friesen (Ekman, 1964, 1965; Ekman & Friesen, 1968, 1969; Ekman, in press) distinguished two methodological approaches for studying nonverbal behavior, namely, measuring judgments about one or another message and measuring the sign vehicles that convey the message.[2] Often either method can be used to answer a question. Take, for example, the question whether facial expressions vary with psychopathology. Suppose a sample was available of facial behavior during interviews with patients who had a diagnosis of schizophrenia or depression, and with a control group who had no psychiatric problems. To utilize the *message judgment* approach, the facial movements in these interviews would be shown to a group of observers, who would be asked whether each person they viewed was normal, schizophrenic, or depressive. If the judgments were accurate, this

would answer the question, showing that facial expressions do convey messages about psychopathology. To utilize the *measurement of sign vehicles* approach, some or all of the facial movements would be classified or counted in some fashion. If the findings showed, for example, that depressives raised the inner corners of their eyebrows more than the other two groups, whereas schizophrenics showed facial movements that very slowly faded off the face, this would also answer the question affirmatively.

Although both approaches can answer the same question, each provides different information. The message judgment approach would show that people can tell from viewing a face whether a person is schizophrenic, depressive, or normal. That cannot be learned from the other approach, which does not determine whether observers can accurately judge this message. But by measuring the sign vehicles it is possible to find out exactly what differs in the faces of the two groups: Is it the timing or the particular movements, or both, that show whether a person is depressive or schizophrenic? That cannot be learned from the first approach, which never determines exactly what the observers respond to when making their judgments.

Let us turn now to some of the other relationships between the outcomes of these two approaches. Consider these cases:

1. Negative findings with message judgment and positive findings with sign vehicle measurement. This suggests that people (at least those used in the study) do not know what to look for or cannot see the differences in facial behavior. Careful measurement of the facial sign vehicles might have revealed hitherto unknown differences. Once known, these clues to psychopathology might make it possible for observers to make judgments accurately. Or perhaps the clues are such that people will never be able to make this judgment accurately when viewing the behavior at real time. The differences in facial behavior might be too subtle to be seen without repeated or slowed viewing and precise measurement.

2. Positive findings with message judgment and negative findings with sign vehicle measurement. The positive results show that there must be some difference in the facial sign vehicles, for how else would the observers achieve accuracy in their judgment? This outcome shows that something must be faulty in the measurement of the sign vehicles. Either the measurement was not reliable or it was selective rather than comprehensive, and there was bad luck in selecting just those facial movements that did not differ.

3. Negative findings with message judgment and negative findings with sign vehicle measurement. This all-too-frequent outcome may occur because the face simply does not provide information about the topic being studied. Or something may have been faulty in the sampling. For example, there may not have been sufficient care in obtaining high agreement among experts about the diagnosis of the patients. Or perhaps the patients were receiving medications that suppressed some behavioral differences. Also, this outcome does not eliminate the possibility that there were differences in facial movement related to psychopathology that the observers did not know about or could not see (thus the message judgment approach failed), and that were missed by a faulty technique for measuring the facial sign vehicle. Was the measurement of sign vehicles comprehensive rather than selective? If it was selective, the possibility always remains that movements unrelated to psychopathology were measured.

The difference between these two approaches – message judgment and the measurement of sign vehicle – has sometimes been confusing, because both may involve observers (Rosenthal, for example, concluded there is little difference between the two; see Section 6.1). It is what the observers do that matters. In message judgment they make *inferences* about something underlying the behavior – emotion, mood, traits, attitudes, personality, and the like. In measuring sign vehicles the observers *describe* the surface of behavior; they count how many times the face moves, or how long a movement lasts, or whether it was a movement of the frontalis or corrugator muscle. (Describing which muscle produced a movement may require an inference, but it is an inference about a physical characteristic, not about underlying psychological phenomena.) Observers who describe behavior are supposed to function like machines, and indeed might someday be replaced by an optical scanner. Techniques for measuring sign vehicles that fail to remove inferences about meaning from the description of behavior will be faulted in the evaluation that follows.

It is not accident that in message judgment studies observers typically are shown a sample of facial behavior at real time, because the purpose usually is to generalize to more natural interpersonal perception. (An exception is the use of still photographs in message judgment studies. These experiments cannot claim any relevance to usual life circumstances.) In sign vehicle measurement there is usually repeated and slowed-motion viewing, because the object is precise description, not observation under natural circumstances.

Though the two approaches can both 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, although there may be study of different types of faces, as in the psychopathology example. In message judgment studies the focus is instead on the person observing the face and/or on the message obtained. Questions have to do with whether a difference is detectable or accurate; there are individual differences among observers, reflecting skill, gender, personality, and the like; messages obtained are best represented as dimensions or categories; and so on.

Facial sign vehicles are measured when the focus is upon 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 young 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 (see 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 if there are any actions that consistently accompany any emphasis (see Ekman, 1980).
3. Is there a difference in the smile during enjoyment as compared to 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 (see Ekman, Friesen, & Ancoli, 1980).
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 (see Ancoli, Kamiya, & Ekman, 1980; Malmstrom, Ekman, & Friesen, 1972).

These 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 requiring 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. But there have been only a few such studies compared to 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 sample 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.

Perhaps the most important obstacle to research measuring sign vehicles has been the lack of any accepted, standard, ready-for-use technique for measuring facial movement. Each investigator who has measured facial movement has invented his technique in large part de novo, rarely making use of the work of his predecessors. Some have seemed to be uninformed by the previous literature. Even the more scholarly have found it difficult to build upon the methods previously reported, because descriptions of facial activity are often vaguer than they appear upon 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.

Coverage

The 14 techniques for measuring facial actions reviewed in this chapter cover a span of 55 years, from Landis's 1924 report to the study by Izard that became available from the author in late 1979. Five were not presented by the authors as methods that could be used by others, but were reported in the course of describing substantive results. They have been included for various reasons. Landis is included because he was among the first to build a measurement system based on the anatomy of muscle action, and his negative findings were influential for the next forty years. Frois-Wittmann (1930) and Fulcher (1942) were both innovative for their times, but their methods and findings have been largely forgotten by the current generation of researchers. McGrew's (1972) behavioral checklist has influenced those studying children from an ethological viewpoint. Nystrom (1974) has been included because there is much interest today in measuring facial action in infants. The other 9 techniques reviewed represent all of the systems for measuring facial movement that have been proposed, some of which have attracted considerable interest and research activity.

A few reports describing facial actions in detail have been omitted. Discussions of facial behavior that did not report a procedure for measurement – such as Hjorstjo (1970), Lightoller (1925), and Seaford (1976), all of which provided very enlightening discussions of the anatomical basis of facial movement – are not included. Depictions of facial expressions primarily designed to train observers to recognize

emotion, rather than measure facial movement (Ekman & Friesen, 1975, are excluded even though some investigators have used them to measure the face (Hiatt, Campos, & Emde, 1977). Izard's Affex (previously called FESM) has also been excluded because observers are required to judge emotion rather than describe the appearance of facial movement. Unlike most message judgment approaches to the measurement of the face, Izard's Affex provides the observers with training about the various clues believed to signal each emotion. There is no way to know, of course, what clues the observers actually rely upon when they make their emotion judgments, because all the investigator obtains is the end point in the observers' inferences. Though the aim of Affex is to provide quick data about emotions, it cannot allow investigation of what indeed are the facial clues to each emotion. Other techniques designed to provide economical measures of emotion (Ekman & Friesen's EMFACS and Izard's MAX) are considered in this chapter because they involve describing facial appearance rather than making direct inferences about underlying states. Reports that used but did not add new methodological features to one of the techniques here reviewed are excluded. Also omitted (except for a later discussion of electromyography [EMG]) are techniques that intrude by attaching something to the subject's face, marking the subject's face, or moving the subject around in front of a camera (Rubenstein, 1969).

The measurement techniques that are reviewed share the features of being unobtrusive; requiring a permanent visual record (video or cinema) that allows slowed or multiple viewing, rather than being applicable to behavior as it occurs; and relying upon an observer who scores or codes behavior according to a set of predetermined categories or items.

This chapter cannot teach the reader how to measure facial actions. Nor does it fully describe most of the measurement techniques, many of which would require a whole chapter, and some an entire book. (Exceptions are the techniques of Birdwhistell, Landis, and Nystrom, each of whom provided little more detail than what is reported here.) Instead, the emphasis is upon the criteria to be considered in evaluating any measurement technique, either one of those available or one that the reader might devise. The strengths and weaknesses of each technique will be made evident, so that the reader is better able to choose which might be best for a particular research problem. Already (Scherer & Ekman, Section 1.4) some mention has been made of the need for reliability and the virtues of a comprehensive measurement system. We

will begin, however, with a different criterion, one that is at the heart of each system: How was it discovered? What basis did the investigator have for proposing his or her technique?

2.2. The basis for deriving units

Each of the 14 measurement techniques contains a list of facial actions such as a brow raise, nose wrinkle, lip corners down, and so on. Measurement includes noting whether any action (or, with some techniques, combination of actions) is present. Later we will consider how each technique describes actions and differentiates one action from another, but here we are concerned with the question how the author decided upon his or her particular list. The lists vary in number of items from a low of 22 to a high of 77. Some actions appear in all techniques, other actions in only some techniques, and still others in just one technique. Sometimes behavior that is treated as a single action by one technique appears subdivided as two distinct actions by others. For example, raising the eyebrows is treated as one behavioral unit by some techniques, but appears as three separate units – inner brow raise, outer brow raise, and the combination of inner and outer brow raise – in other techniques. Most authors did not explain what they considered when they included or excluded a facial action, what basis they had for subdividing a unit another researcher had treated as a single action, or why they found it wise to collapse a distinction drawn by another investigator. In fact, most did not acknowledge the work of their predecessors, but instead acted as if they had invented their system and had no knowledge of differences between it and the systems of their earlier or contemporary colleagues.[3]

Investigators – often failing to specify the sample, setting, or persons viewed – usually said only that they looked at behavior and that their list of facial actions was simply the product of what they saw. Something more is needed, however, to account for the differences among these techniques, even allowing for the fact that each investigator observed a different behavior sample. What stood out, which attributes were noticed when an action occurred, how the flow of behavior was segmented by the investigator probably depended upon theoretical commitments. Only a few were explicit.

Birdwhistell (1952) tried to organize units and select behavior to construct a system paralleling linguistic units.[4] Grant (1969) advocated the selection and organization of measurement units according to function. This puts the cart before the horse, because the measurement

technique so constructed was to be used to discover the function of those very behaviors. Among ethologists, Blurton Jones (1971) was most explicit in considering the anatomical basis for facial actions, although he did not say that this was the final or even the major basis for his decisions about what to include, and he did not specify how he arrived at his list of minimal units of behavior.

Ekman, Friesen, and Tomkins (1971), in contrast to the aforementioned investigators, derived their list of facial actions from explicit theory about the facial actions relevant to emotion, rather than from observation of some sample of behavior. The cart-before-horse criticism applies to them also. Although they could find out whether the actions proposed for one emotion do or do not accurately reflect that emotion, they could not discover signals for emotion that they did not know about in advance. Izard, eight years later, also used theory about emotion signals as the basis for selecting actions to score in his measurement technique MAX. His decisions were based on inspection of still photographs of posed emotions that had yielded high agreement among observers who made global judgments about emotion.[5]

The anatomical basis of facial action provided a third totally different basis for deriving units of behavior. The measurement units were presumably based on what the muscles allow the face to do. Because we all have the same muscles (for all practical purposes), this approach might be expected to have led the investigators who followed it to arrive at the same listings of facial actions. This is not the case. For example, Landis had 22 actions and Frois-Wittmann 28, and yet they both claimed to have based their measurement units on the anatomy of facial action. In part, the discrepancies occurred because of explicit decisions to select only certain actions. Most standard anatomy texts list many, usually not all, facial muscles with rather simple, only partially correct, and usually quite incomplete accounts of how each muscle changes appearance. Most investigators who based their technique on anatomy selected only some muscles, and usually did not explain the basis for their selection. Ekman and Friesen (1976, 1978) and Ermiane and Gergerian (1978) were exceptions, each attempting to determine all the actions the anatomy allows. Both studies attempted to determine this by systematically exploring the activity of each single muscle; Ekman and Friesen also resurrected Duchenne's (1862) technique of determining how muscles change appearance by inserting a needle into and electrically stimulating muscles.

The discrepancies among the most recent techniques (Ekman & Friesen; Ermiane & Gergerian; Izard's MAX) are due to differences in

purpose and in procedure for obtaining reliability. Both Ekman and Friesen and Ermiane and Gergerian attempted to include in their lists changes in appearance that are independent of each other. If a muscle contraction would produce two or three changes in appearance, these were gathered together as multiple indexes of the activity of one unit or muscle. For example, when the entire frontalis muscle acts, it will (1) raise the eyebrows; (2) produce horizontal furrows running across the forehead (except in infants, who have a fatty pad in the forehead blocking such wrinkles); and (3) expose more of the eye cover fold (the skin between the upper eyelid and the eyebrow). Both Ekman and Friesen and Ermiane and Gergerian listed these multiple signs together as different ways of recognizing that this one action had occurred. Izard, however, treated signs (1) and (2) of frontalis muscle activity as separate measurement units, giving each equal, independent, separate status, failing to recognize that they are signs of the same action. He ignored sign (3).

Izard also differed from the others in selecting only movements that he judged relevant to emotion. Ekman and Friesen and Ermiane and Gergerian intended to include all the possible appearance changes that the muscles can produce. This sometimes meant creating more than one measurement unit, if use of different strands of a single muscle or different portions of that muscle was found to produce visibly different changes in appearance. For example, Ekman and Friesen and Ermiane and Gergerian distinguished a number of different facial action units that are based on various uses of what anatomists have termed one muscle – the orbicularis oris, which circles the lips. Izard included only some of these separate appearance changes. Strangely, Izard excluded specific actions that are said by many theorists to signal emotions and that are shown by Ekman and Friesen's data to be emotion signals. Izard and Dougherty (1981) say that actions were dropped that were not efficient, but inspection of that article and of earlier versions of Izard's scoring technique (FMCS) suggests instead that Izard never considered a number of facial actions important to differentiating among emotions.

The Ekman and Friesen technique differed from the others in another important respect. Anatomy was only part of their basis for the derivation of measurement units. They also determined whether observers could reliably distinguish all of the 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. If Ekman and Friesen erred, it was on the side of caution, by excluding distinctions that observers with considerable training

might perhaps be able to distinguish. The opposite error may have been made by Ermaine and Gergerian and by Izard. They included distinctions without reporting exploration or test of whether each and every distinction could reliably be made by those who learn their system (see Section 2.7 on reliability).

Tables 2.1, 2.2, and 2.3 and the chapter appendix compare the 14 measurement techniques on each of the criteria (arranged as columns) that are discussed. The basis for deriving units provides the order in which the 14 techniques appear: first the one system that is linguistically based; then those that are ethologically based; then one that is theoretically based; and finally those based on the anatomy of facial action.

2.3. Comprehensiveness or selectivity

Three aspects of facial movement can be measured either selectively or comprehensively. Most investigators have considered how to measure only the type of action, not its intensity or its timing. *Type* refers to whether it was a brow raise, or an inner corner brow raise, or a brow lower, or some other action. *Intensity* refers to the magnitude of the appearance change resulting from any single facial action. *Timing* refers to the duration of the movement, whether it was abrupt or gradual in onset, and so on.

Type of action

A technique for measuring the type of facial action can be selective, measuring only some of the actions that can occur, or it may claim to be comprehensive, providing a means of measuring all visible facial action. There are advantages and disadvantages in each case. If the technique is selective, it is important to know what has been excluded; and if it claims to be comprehensive, there must be some evidence to establish that this is indeed the case.

The great advantage of a selective technique is economy. Because only some of the mass of facial actions must be attended to, the work can be done more quickly. Suppose an investigator wants to measure whether fear is reduced by exposure to one set of instructions versus another. A measurement technique that allows measurement of just the occurrence of three or four fear facial expressions would be ideal, because it will not matter if the occurrence of anger, disgust, distress, or some other emotion is missed. Even if the technique does not include *all* of the fear facial expressions (and at this time there is no conclusive or even

Table 2.1. *Summary of methods for measuring facial behavior for units and comprehensiveness*

		Comprehensiveness		
	Basis for deriving units	Type of action	Intensity of action	Timing of action
Linguistically based				
Birdwhistell (1952)	Observation of interpersonal behavior; parallel linguistic units	Not claimed to be comprehensive; 53 actions	No provision	No provision
Ethologically based				
Blurton Jones (1971)	Observation of 500 still photographs of 2–5-year-old children	Measures any child's facial expressions; 52 actions	6 degrees of eye openness; 4 degrees of lip separation; 2 degrees of frowns	No provision
Brannigan & Humphries (1972)	Observation of children and adults	Not claimed to be comprehensive; 70 actions	No provision	No provision
Grant (1969)	Observation of children and adults	Not claimed to be comprehensive; 53 actions	No provision	No provision
McGrew (1972)	Observation of 3–4-year-old children	Not claimed to be comprehensive; 31 actions	No provision	No provision
Nystrom (1974)	Observation of 1-month-old infants	Not claimed to be comprehensive; 35 descriptors	No provision	No provision
Young & Decarie (1977)	Observation of 36 infants	Measures 42 facial configurations; selected only to be relevant to emotion in the last quarter of first year in six test situations	No provision	No provision
Theoretically based				
Ekman, Friesen, & Tomkins (1971)	Theory about emotion expression	Measures signs of just 6 emotions; 77 descrip-	No provision	Start–stop

Nystrom (1974)	Observation of 1-month-old infants	Not claimed to be comprehensive; 35 descriptors	No provision	No provision
Young & Decarie (1977)	Observation of 36 infants	Measures 42 facial configurations; selected only to be relevant to emotion in the last quarter of first year in six test situations	No provision	No provision
Theoretically based				
Ekman, Friesen, & Tomkins (1971)	Theory about emotion expression	Measures signs of just 6 emotions; 77 descriptors	No provision	Start–stop
Izard MAX (1979[b])	Theory about emotion signals; data from posed still photographs	Measures just actions needed to identify emotion in infants; 27 descriptors	No provision	Start–stop
Anatomically based				
Ekman & Friesen (1976, 1978)	Muscular	Measures all visible movements; 44 action units that singly or in combination can score any observed action	Four actions have 3-point rating on intensity; provision to rate intensity in other actions	Start–stop and onset–apex–offset
Frois-Wittmann (1930)	Muscular	Not claimed to be comprehensive; 28 descriptors	No provision	No provision
Fulcher (1942)	Muscular	Not claimed to be comprehensive; absence/presence of 16 muscular actions	Amount of movement in each of three facial areas rated	No provision
Ermiane & Gergerian (1978)	Muscular	Measures all visible movements; 27 muscle actions	Each action rated on 3-point intensity scale	No provision
Landis (1924)	Muscular	Not claimed to be comprehensive; 22 descriptors	Each action rated on 4-point intensity scale	No provision

definitive evidence about all the facial actions for *any* emotion), a selective technique could be useful. It might not matter that some or even most fear expressions were not scored, nor that blends of fear with other emotions were not scored; enough might be measured to show the effect. If the findings were negative, however, the investigator would not know whether the cause was an inadequate experimental treatment (in this example, the instructions might not have differed sufficiently) or failure to measure all of the fear expressions. In such an instance the investigator might want to turn to a comprehensive technique.

Some questions require a comprehensive technique and cannot be answered with a selective one. Suppose the investigator wishes to discover which facial actions signal fear, anger, sadness, and so on.[6] Or perhaps he or she wishes to discover whether different actions are employed to serve a linguistic rather than an emotive function, or to learn what people show on their faces when their heart rates show a sharp acceleration, or whether there are cultural or social class differences in facial actions during a greeting. A comprehensive technique would have to be employed. Once there was reasonably conclusive evidence on any of these issues, then such evidence could provide the basis for selective use of portions of a comprehensive system. For example, Ekman, Friesen, and Simons (in preparation), building upon the earlier research of Landis and Hunt (1939), have strong evidence about the particular combination of facial actions and the timing of those actions that index the startle reaction.[7] Once that has been replicated by other laboratories, those interested in the startle in particular could utilize just that portion of Ekman and Friesen's comprehensive scoring technique.

Only a comprehensive technique allows for discovery of actions that the investigator did not know about in advance and permits a complete test of an a priori theory about facial sign vehicles. A third advantage of a comprehensive technique is that it provides a common nomenclature for descriptions of facial behavior. If many investigators were to use the same comprehensive technique, comparison of findings would be facilitated because investigators, even those who used it selectively, would key their units to a single list of facial actions. Investigators considering selective scoring might well want first to study a comprehensive technique, in order to become acquainted with the entire array of facial actions, so that they could be explicit about what it is they are choosing not to measure.

Wedded to these advantages of comprehensive facial scoring is the disadvantage of cost. It takes more time to learn a comprehensive technique, and it takes more time to apply it, for nothing (presumably) is left out.

It is no
comprehensi
anatomically
comprehensi
behavior wou
achieving cor
comprehensi
muscles prod
first seem, be
expressions a
Even the smi
major muscle
not every sm
happens to a
sum of the ch
activity of or
important, the
ways of recog
ways of scorin
combine in si
Friesen techni

A last issue
type of facial ac
system is what
of facial move
part of the ur
comprehensive
actions. An er
niques claimin
and Gergerian
females of dive
grounds, in a v
been used this

Alternatively
mentally gener
and Friesen e
producing volu
combinations o
tions of the acti
possible combi
they believe th
application to d

It is no accident that the only techniques that claim to be comprehensive – Ekman and Friesen and Ermiane and Gergerian – were anatomically based. An inductive approach would be too costly if comprehensiveness was the goal. Too large a sample of diversified behavior would have to be observed to have a reasonable likelihood of achieving completeness. By contrast, it should be possible to achieve comprehensiveness by exploring how each muscle works, because the muscles produce the actions observed. This is not as simple as it might first seem, because muscles can act in concert, not just singly. Facial expressions are rarely the consequence of the activity of a single muscle. Even the smile, which is principally the work of the single zygomatic major muscle, typically involves two or three other muscles as well, and not every smile involves the same other muscles. Moreover, what happens to appearance when muscles act in concert is not always the sum of the changes associated with each of the components. And the activity of one muscle may obscure the presence of another. It is important, therefore, that a comprehensive technique list not simply the ways of recognizing how each single facial action appears, but also the ways of scoring the occurrence of these units of facial action when they combine in simultaneous or overlapping time. Only the Ekman and Friesen technique has done so.

A last issue regarding how comprehensively a technique measures the *type* of facial action is what evidence is provided to demonstrate that the system is what it claims to be. One wants to know whether the universe of facial movement can be described by the technique, or at least what part of the universe has been omitted. If there is uncertainty about comprehensiveness it should be clear whether it is about just some or all actions. An empirical answer would be possible if either of the techniques claiming comprehensiveness (Ekman and Friesen and Ermiane and Gergerian) had scored large samples of facial actions of males and females of diverse ages, from various cultural, ethnic, and class backgrounds, in a wide variety of social and individual settings. Neither has been used this extensively.

Alternatively, comprehensiveness could be determined by experimentally generating all possible permutations of facial actions. Ekman and Friesen explored the comprehensiveness of their technique by producing voluntarily on their own faces more than 7,000 different combinations of facial muscular actions. These included all permutations of the actions in the forehead area, and for the lower face all of the possible combinations of two muscles and of three muscles. Although they believe their system is relatively comprehensive[8] only time and application to diverse samples of facial behavior will establish it to be so.

Ermiane and Gergerian did not provide any evidence of comprehensiveness. They determined only that their system would describe the actions of single muscles, and a few of the combined actions of two or three muscles.

Intensity of action

Actions vary not only in type (inner corner brow raise versus raise of the entire brow) but also in intensity. A brow raise may be weak or strong; the lift of the brow, the extent of exposure of the eye cover fold and gathering of skin on the forehead, may be very slight or great. The intensity of a facial action may be of interest for a variety of reasons. For example, Ekman et al. (1980) found that the intensity of zygomatic major muscle action was correlated with retrospective self-reports about the intensity of happiness experienced.[9]

Ermiane and Gergerian was the only one of the 13 other techniques to provide for comprehensive measurement of intensity. Nine of the techniques treated facial action as an all-or-nothing phenomenon, or as if there were evidence that variations in intensity are without significance. One (Grant) even confused intensity with type of action, listing as different action types appearance changes that are due only to variations in intensity. A few made provision for scoring the intensity of four or five actions (see Table 2.1). In recent unpublished work, Ekman and Friesen found that the logic provided in their scoring system for measuring the intensity of four actions can be extended to the other facial actions, but evidence has not yet been provided that such extensions can be made reliably for all the actions in their technique.

Timing of action

A facial action has a starting and a stopping point. It is often more difficult to ascertain the exact determination of these points than to decide which action occurred (see the discussion of timing in Section 2.6). From start to stop, other aspects of timing may be distinguished:

1. *Onset time*: the length of time from the start until the movement reaches a plateau where no further increase in muscular action can be observed
2. *Apex time*: the duration of that plateau
3. *Offset time*: the length of time from the end of the apex to the point where the muscle is no longer acting

Onsets and offsets may vary not only in duration but in smoothness; for example, an offset may decline at a steady rate, or steps may be

apparent. Similarly, an apex may be steady or there may be noticeable fluctuations in intensity before the offset begins. When examined closely the separate actions that compose a facial expression do not start, reach an apex, and stop simultaneously. In even a common expression, such as surprise, the raising of the eyebrows may reach an apex while the dropping of the jaw is still in onset.

For some questions it is possible that simple counts of the occurrence of particular actions may be sufficient, without measurements of onset, apex, and offset. The investigator may want to know only how often or for how long a person raised the brow, wrinkled the nose, or depressed the lip corners. Even when interest is limited to simple summary measures of the occurrence of single actions, there is no rationale for using frequency rather than duration measures (which require stop–start determination) other than economy. A frequency count will underrepresent those actions which go on for long periods of time and overrepresent frequent brief actions.

Limiting measurement to single actions is hazardous regardless of whether frequency or duration is measured. Nose wrinkling, for example, may signify one thing when it occurs in overlapping time with a lower lip depression (disgust) and something quite different when it flashes momentarily while the lip corners are pulled upwards (an action that Ekman and Friesen suggest functions like a wink to accentuate a smile). A pulling down of the lip corners may signify sadness when it accompanies raised inner corners of the brows with drooping upper eyelids. When this same action occurs with the entire brow raised and the lower lip pushed up it may be a disbelief gesture. These interpretations, which have not all been tested, cannot be tested unless the timing of actions is measured. What evidence does exist (Ekman & Friesen, 1978) suggests that it is unwise to measure the face as if each action can be counted separately, as if each action has an invariant meaning apart from other actions that overlap in time.

Measurement of combinations of facial actions (what is usually meant by an expression) requires at least a determination that actions overlap, if not precise determination of the stopping and starting points of each action. Ekman and Friesen (1978) further suggest that it is overlap in the apex that is crucial to determining whether actions that co-occur are organized as part of the same event, signal, or expression. Their reasoning is that when one action begins (onset) while another action is fading (offset), it is not likely that they have been centrally directed as part of the same signal. Suppose, for example, that there has been an overlap in the apex of brow lowering, tightening and pressing together

of the red parts of the lips, and raising the upper eyelid. Ekman and Friesen have hypothesized that these elements compose one of the anger expressions. Overlap in the apex of these actions would support their notion that an anger signal had occurred and that these actions should be so counted, and not tallied separately. Let us suppose that there was also a nose wrinkle, with an apex overlapping these anger actions. Ekman and Friesen suggest that this would be a blend of disgust with anger. If the nose wrinkling reached its apex as these anger actions were in offset, they suggest that it be characterized as a sequence of anger followed by disgust. Test of these hypotheses requires precise measurement of onset, apex, and offset.

A number of other research questions also require comprehensive measurement of the timing of facial actions. For example, does a brow raise and upper eyelid raise occur before or during an increase in loudness in speech or a deceleration in heart rate? Ekman, Friesen, and Simons (in preparation) have found that onset time is crucial in isolating from idiosyncratic facial actions those muscular actions which always occur in unanticipated startle reactions. Only actions that began within 0.1 second were evident in all unanticipated startles; offset time did not distinguish the idiosyncratic from uniform facial actions. In another situation offset time, rather than onset, may be crucial; for example, Ekman and Friesen (1975, chap. 11) hypothesized that stepped offsets occur more often in deceptive than in felt emotional expressions.

Most of the 14 techniques do not describe procedures for measuring starting and stopping points and totally ignore onset, offset and apex measurement. The data reported usually consists only of frequency counts. Ekman and Friesen's technique is the only one to describe how to measure these different aspects of timing. In a study now in progress these authors are comparing the relative validity scores of such comprehensive measurements of timing with a more economical frequency checklist version of their Facial Action Coding System (see Section 2.9).

2.4. Depicting facial measurement units

It is not as easy as it may at first seem to depict clearly what is referred to by a facial measurement unit. Some authors did not bother, because they did not expect others to try to use their methods. Regrettably, this lack of clarity also has caused some uncertainty about their substantive results. Take the example "down corners mouth," which is found in the measurement techniques of Birdwhistell, Brannigan and Humphries

(1972), Grant, and Nystrom. Does this phrase describe instances in which the mouth corners have been pulled down? Or those in which the mouth corners are down because the chin and lower lip have been pushed up in the middle? Or does it refer just to expressions in which the mouth corners are down because the center of the upper lip has been raised? Or is it all of them?

The first column in Table 2.2 describes how measurements were depicted in each of the 14 techniques. The chapter appendix lists how a particular facial action (brow raise) was depicted by each technique.

Most techniques used but a few words to describe each measurement unit. Some supplemented this description with a few still photographs. Only three techniques went beyond this step to provide more thorough illustration of each unit. Ekman and Friesen, Ermiane and Gergerian, and Izard's MAX technique all provided visual illustrations of every measurement unit. All provided some explanations of the anatomical basis of each action, Ekman and Friesen and Ermiane and Gergerian more thoroughly than Izard. Ermiane and Gergerian provided still photographs of each action and combination considered; Izard provided video, photographs, and drawings; and Ekman and Friesen provided still photographs, cinema, and video illustrations.

2.5. Separating inference from description

Although many investigators have been interested in inferring something about the signal value or function of facial actions, not all have recognized that such inferences should not be intermixed with descriptions in their measurement techniques. The measurement must be made in noninferential terms that describe the behavior, so that inferences about underlying states, antecedent events, or consequent actions can be tested by empirical evidence.

Mixing inference with description may also make the measurements quite misleading. Few single-muscle actions have an invariant meaning. Take the example of the so-called frown (lowering and drawing the brows together). This action is not always a sign of negative affect; depending upon the timing of the action, what other actions co-occur with it, and the situational context, it may signify quite different matters. It would be misleading to be identifying the occurrence of a frown when the brow lowering is signaling concentration, or conversational emphasis.

Because humans make the measurement, inferences cannot be elimi-

Table 2.2. *Summary of methods for measuring facial behavior: unit depiction, inference/description, and application*

	Way in which each unit is depicted	Use of inference or description	Types of records and persons to which measurement has been applied
Linguistically based			
Birdwhistell (1952)	Two or three words	Mixed: e.g., *pout, smile, sneer*	Not known
Ethologically based			
Blurton Jones (1971)	Verbal description of changed appearance of features, a few drawings and illustrative photos	Mostly description but a few inferential terms: e.g., *frown, pout*	Infants and children
Brannigan & Humphries (1972)	Verbal description	Mixed: e.g., *wry smile, angry frown, sad frown, threat*	Children and adults
Grant (1969)	Primarily verbal description, some photos	Mixed: *sad frown, aggressive frown, smile, sneer,* etc	Children and adults
McGrew (1972)	Verbal description; compared to Grant, Blurton Jones	Mostly description but a few inferential terms: e.g., *pout, frown, grin*	Children
Nystrom (1974)	Verbal description	Description	Neonates
Young & Decarie (1977)	Verbal description	Mixed: *fear face, sad face, shy smile,* etc.	Infants in last quarter of first year
Theoretically based			
Ekman, Friesen, & Tomkins (1971)	Photographs of descriptor	Description	Video and still photos of adults' posed and spontaneous expressions
Izard MAX (1979[b])	Verbal description, photos, drawings, and video	Description	Video of infants

Table 2.2. *(cont.)*

	Way in which each unit is depicted	Use of inference or description	Types of records and persons to which measurement has been applied
Anatomically based			
Ekman & Friesen (1976, 1978)	Verbal description, still photos, and cinema illustrations of each action and certain combinations of actions	Description	Spontaneous, deliberate, and posed video and photos of neonates, children, adults, deaf stutterers, mental patients
Frois-Wittmann (1930)	Verbal description; very brief	Only one inferential term: *frown*	Still photos of poses by one adult
Fulcher (1942)	Verbal description; very brief	Description	Films of poses by blind and sighted children
Ermiane & Gergerian (1978)	Verbal description, still photos	Description	Adult poses and patients' spontaneous photographs
Landis (1924)	Verbal description	Description	Neonates

nated, but they need not be encouraged or required. If the person scoring a face identifies the brows being lowered and/or drawn together, the scorer may still make the inference that he or she is describing a frown. But Ekman and Friesen (1978) reported that when people use a measurement technique that is solely descriptive, as time passes the scorer increasingly focuses on the behavioral discriminations and is rarely aware of the possible meaning of the behavior. Although there can be no guarantee that inferences are not being drawn, a measurement technique should neither encourage nor require inferences about meaning by the terminology or descriptions it employs.

Both Ekman and Friesen and Izard separated their hypotheses about the signal value of facial actions from the descriptive materials to be used

in training a person to measure facial behavior. Ermiane and Gergerian intermixed inferences about the meaning of behavior with the information necessary to learn their descriptive system. Theirs is the only technique to contain inferences about how given facial actions are indicative of specific personality processes and types of psychopathology. Birdwhistell, Blurton Jones, Brannigan and Humphries, Grant, McGrew, Young and Decarie (1977), and Frois-Wittmann all used some inferential or emotional terms (e.g., *frown, smile, sneer, angry frown*) mixed in with descriptive terms. (This is not always evident from the chapter appendix, because not all who mixed inference with description did so for the brow raise.)

Both Ekman and Friesen and Izard listed hypotheses about the emotion signaled by particular facial actions. Ekman and Friesen were explicit about the particular combinations of units they considered as emotion signals; 1,000 such predictions were included in their published system, and more than 2,000 more are contained in a forthcoming report (Friesen & Ekman, in preparation). Izard's MAX contains only those facial actions which, he claims, distinguish among the emotions. Ekman and Friesen have evidence that Izard is wrong, that he has excluded a number of actions relevant to emotions. For example, Izard does not include levator labii superioris caput infraorbitalis, an action relevant to both disgust and anger, except when this muscle acts unilaterally. Ekman, Friesen, and Ancoli (1980) found that bilateral evidence of this muscle correlated with the subjective report of disgust. Ekman and Friesen also found that when this action is accompanied by the narrowing of the red margins of the lips (another action ignored by Izard), the signal changes from disgust to anger. These errors are the product of limited sampling: Izard chose his actions on the basis of what he observed in a set of photographs of posed emotions.

2.6. Types of records and persons to which measurement has been applied

Still or motion records

Although a number of techniques claim that they can be used with motion records, most have not dealt with the complexities in the timing of facial action that a motion record reveals. These investigators may never have been confronted with the complexity of the temporal organization of facial actions because of either the type of behavior or

the type of record they examined. If only posed expressions were measured (as in the case of Ermiane & Gergerian), variations in timing might not be apparent. Posers generally try to perform all the required movements at once, in overlapping time, with similar very short onsets, long-held apexes, and abrupt short offsets. Variations would not be apparent, nor would the reason to measure them. An investigator who used his or her method only to score still photographs also might not know of these complexities in timing, because the camera shutter freezes all action. Though Izard has scored some motion records, he preselected only certain brief segments of videotape to score, segments in which the infants seemed to be emitting expressions that looked like those in posed photographs of adults. Thus he has not dealt with the complexities that a motion record reveals. Other investigators may have failed to consider the timing of facial movement because they tried to apply their systems in real time, as the behavior occurred, and even if they had videotape or film, they may not have examined the records in slowed or repeated replay.

It will be most important for investigators to make use of motion measuring the timing of facial actions whenever they want to study spontaneous behavior, taking a strictly descriptive approach; or interrelate facial activity and some other simultaneous behavior (speech, respiration, body movement, etc.); or distinguish configurations in which the temporary organization of multiple facial actions suggests that they be considered parts of the same signal or expression. (See the discussion of the research questions that require measurement of timing in Section 2.3.)

Modifications for varying age levels

Ideally, a facial measurement system should be applicable to the study of individuals of any age, by making provision for any modifications needed to measure infants or the aged. The appearance of certain facial actions is quite different in neonates and infants from what it is in young children and adults. Oster (1978; Oster & Ekman, 1978), who worked with Ekman and Friesen during the final stages in the development of their measurement system, has studied the neuroanatomical basis for these differences. She has provided (Oster & Rosenstein, in preparation) a set of transformations for utilizing the Ekman and Friesen system with neonates and infants. Izard's MAX technique is specifically limited to measuring infants, but he provides only a few overly general descrip-

tions of potentially confusing infant–adult differences. No other investigator has attended to this problem, not even those who measured young infants and neonates.

Parallel problems may occur in measuring facial activity in quite elderly people, because age signs may necessitate some modifications in scoring rules to avoid mistakes in identifying certain actions. No one has considered this.

2.7. Reliability

The need for reliability is obvious to psychologists. To some anthropologists and sociologists the quest for reliability has seemed a peculiar madness that deflects psychologists from the real problem at hand. For example, Margaret Mead, in the last years of her life, wrote "Psychologists . . . are more interested in validity and reliability than in what they are actually studying" (1975, p. 211). Yet if a measurement system cannot be shown to be reliable, there is no way of knowing whether even the investigator who invented the system recognizes the same facial action when it twice occurs. The need to demonstrate reliability seems especially important with facial behavior. For here there is an enormous variety of behaviors that can occur, with no names for most. And those who have observed facial actions have produced very different catalogs.

Some ethologists (e.g., Young & Decarie) have argued that if the same finding is obtained in two independent studies, there is no need to demonstrate that the measurement technique was reliable. This reasoning should not be applied to the area of facial measurement, where there have been completely contradictory reports by different investigators (e.g., the argument about universality between Birdwhistell and Ekman). If we knew that Birdwhistell and Ekman had each used a reliable measurement technique (preferably the same one), at least we could be certain about what was seen, and search differences in sampling, situation, or interpretation as sources of their disagreement. When a measurement technique is intended to be usable by other investigators, it is especially important for its originator to demonstrate that he or she as well as others can use it reliably. (See also Section 2.1, where reliability was discussed in the context of the relationship between the outcomes of message judgment studies and measurement of sign vehicle studies.)

Let us consider now various aspects of reliability, for it is not a simple matter to establish. A number of requirements can be enumerated:

1. The researcher, rather than just giving an overall index of agreement, should provide data to show that high agreement can be reached about the scoring of specific facial actions. Typically, some actions are easier to recognize than others. Unless reliability data are reported for the scoring of each facial unit, it is not possible to evaluate which discriminations may be less reliable.

2. Data on reliability should be reported from the measurement of spontaneous, not just posed, behavior, and from the flow of behavior as revealed in a motion record, not just from still photographs or slices abstracted from video or cinema, which may yield higher agreement.

3. Reliability data should be provided for (*a*) infants, (*b*) children, (*c*) adults, and (*d*) aged populations, because reliability on just one group does not guarantee reliability on the others.

4. The most common source of unreliability in behavioral measurement, whether it be of face or of body, is the failure of one person to see what another scores. Usually this occurs when an action is small in size. This source of disagreement can be attenuated if the technique specifies a threshold that must be surpassed for the action to be scored. Specifying minimum thresholds alerts the persons doing the scoring to subtle signs and provides explicit bases for decisions about when a change in appearance is likely to be ambiguous. A technique that provides such threshold definitions should therefore yield higher agreement.

5. Reliability should be reported not only for the person(s) who developed the technique, but also for learners who did not previously have experience with facial measurement. Data about the range of reliabilities achieved by new learners should be provided and compared to those for experienced or expert scorers. A technique will be more generally useful if it can be learned independently, without direct instruction from the developer. This usually requires a self-instructional set of materials, practice materials with correct answers, and a final test for the learner to take.

6. Reliability should be reported for the scoring of not just (*a*) the type of action, but also (*b*) the intensity of actions and (*c*) the timing of actions.

Of the 14 measurement techniques, 5 did not report data on any aspect of reliability. The other 9 provided fairly sparse data on reliability, with the exception of Ekman and Friesen and Izard. Even these techniques did not meet all the requirements just listed. Table 2.3 lists the specific reliability requirements met by each technique.

Table 2.3. *Summary of methods for measuring facial behavior: reliability and validity*

	Reliability	Validity			
		Descriptive	Emotional	Conversational	Other
Linguistically based					
Birdwhistell (1952)	Not reported	None	None	None	None
Ethologically based					
Blurton Jones (1971)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	None	None	None
Brannigan & Humphries (1972)	Not reported	None	None	None	None
Grant (1969)	Not reported	None	None	None	Predicts severity of mental illness, but no data reported
McGrew (1972)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	Spontaneous	None	Predicts gender differences & relation to agonistic interaction
Nystrom (1974)	Data reported on requirements 1, 2, 3*b*, 6*a*	None	None	None	None
Young & Decarie (1977)	Not determined by authors	None	Spontaneous, but no data reported	None	Said to differentiate infants' response when mother departs and when she frustrates, but no data reported
Theoretically based					
Ekman, Friesen, & Tomkins (1971)	Data reported on requirements 2 and 3*c*	None	Posed and spontaneous: positive vs. negative, stressful vs. neutral film conditions;	None	Predicts attribution of emotion

Young & Decarie (1977)	Not determined by authors	None	Spontaneous, but no data reported	None	Said to differentiate infants' response when mother departs and when she frustrates, but no data reported
Theoretically based					
Ekman, Friesen, & Tomkins (1971)	Data reported on requirements 2 and 3*c*	None	Posed and spontaneous: positive vs. negative, stressful vs. neutral film conditions; differentiates patterns of heart rate	None	Predicts attribution of emotion
Izard MAX (1979[b])	Data reported on requirements 2, 3*a–b*, 5, 6*a*	None	Posed	None	Provides preliminary data on relations to vocalization and body movement in infants
Anatomically based					
Ekman & Friesen (1976, 1978)	Data reported on requirements 1, 2, 3*a–c*, 4, 5, 6*a* & *c*	Meets performed actions and EMG criteria	Posed and spontaneous; measures intensity and type of emotion; differentiates startle reaction; differentiates certain deliberate from spontaneous expressions	Measures syntactic and emphasis signals	None
Frois-Wittmann (1930)	Not reported	None	Posed	None	Predicts developmental changes; compares blind and sighted

Table 2.3. *(cont.)*

	Reliability	Validity			
		Descriptive	Emotional	Conversational	Other
Fulcher (1942)	Data reported on requirements 2, 3*b*, 6*a*	None	Posed	None	None
Ermiane & Gergerian (1978)	Data reported only on scoring photos of poses and on requirement 3*c*	None	Posed	None	None
Landis (1924)	Not reported	None	None	None	Predicts individual differences

2.8. Validity

Descriptive validity

The validity of a technique designed to measure facial movement entails questions on a number of levels. Most specifically (and concretely), 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, what criterion to utilize independently of the facial measurement technique itself. Two 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. Ekman and Friesen's Facial Action Coding System (FACS) accurately distinguished the actions the performers had been instructed to make.
2. Electrical activity criterion: Ekman and Friesen, in collaboration with Schwartz (Ekman, Schwartz, & Friesen, in preparation), 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. (This study is described in more detail in Section 2.10.)

Utility or validity

Some measurement techniques contain hypotheses about the particular facial actions that signal particular emotions (Ekman and Friesen; Ekman, Friesen, and Tomkins; Ermiane and Gergerian; Izard). For these techniques it is appropriate to ask whether the hypotheses are correct, but the answer does not pertain to the validity of the techniques, only to that of the hypotheses. Suppose the facial behaviors found to signal emotion were exactly the opposite of what had been hypothesized by the developer of the technique. Such evidence would not show that the technique was invalid, only that the hypotheses were wrong. In fact, the discovery that the hypotheses were wrong would itself require that the

technique measure facial movement accurately. Suppose a study not only failed to support the investigator's hypotheses about the actions that signal emotions but found that there were no facial actions related to emotion. If one could discount the possibility that the sample did not include emotional behavior, this might suggest that the facial measurement technique was not *relevant* to emotion. It might have measured just those facial behaviors which are unrelated to emotion. Another technique applied to the same sample of facial behavior might uncover the actions related to emotion.

Two techniques (Ekman and Friesen and Ermiane and Gergerian) claim not to be specific to the measurement of any one type of message, such as emotion, but to be of general utility, suitable for the study of any question for which facial movement must be measured. Such a claim can be evaluated by evidence that the technique has obtained results when studying a number of different matters.

Posed emotions. Many techniques have been shown to be able to differentiate poses of emotion or judgments of emotion poses: Ekman and Friesen; Ekman, Friesen, and Tomkins; Ermiane and Gergerian; Frois-Wittman; Fulcher; Izard. In the studies that used a selective technique it is not possible to know whether there might have been other facial actions not included in the scoring technique that might have predicted the emotion poses or judgments just as well or better. The two comprehensive techniques – Ekman and Friesen and Ermiane and Gergerian – provided that information. They were able to show that it was the movements they specified as emotion-relevant, not other movements, that were signs of particular emotions. Ekman and Friesen's FACS also predicted not only *which* emotion was posed or judged, but the *intensity* of emotion as well.

Poses, however, by definition are artificial. Although they may resemble spontaneous facial expressions in some respects (see Ekman, in press), one difference is that they are likely to be easier to score. The onset may be more coordinated and abrupt, the apex frozen, and the scope very intense or exaggerated (see the discussion in Section 2.6). Evidence that a technique is a valid measure of emotion cannot rest just upon measurement of poses; it is necessary to determine that the measurement will be valid when it measures spontaneous emotional expression.

Spontaneous emotions. A number of studies have shown the validity of Ekman and Friesen's FACS in measuring the occurrence of spontaneous

emotional expressions. Ancoli (Ancoli, 1979; Ancoli et al., 1980) studied autonomic nervous system (ANS) responses when subjects watched a pleasant or stress-inducing film. A different pattern of ANS response during the two films was found only during the times in each film-viewing period when the face registered maximal emotional response. In another study of that data, Ekman et al. (1980) found that FACS accurately predicted the subjects' retrospective reports of their emotional experience while watching the films: the intensity of happy feelings, the intensity of negative feelings, and, specifically, the intensity of the emotion disgust. Ekman, Friesen, and Simons (in preparation) differentiated the specific facial actions that signify a startle reaction from the emotional reactions subsequent to being startled. Both the type of actions and the onset time were crucial to this distinction. They also were able to differentiate a genuine from a simulated startle accurately. Ekman, Hager, and Friesen (1981) examined the differences between deliberate facial movements and spontaneous emotional expressions. Scoring the intensity of each specific facial action on each side of the face, they found that requested facial movements were asymmetrical more often than spontaneous emotional expressions: The actions usually were more intense on the left side of the face for the deliberate, but not for the spontaneous, emotional expressions. Krause (1978) utilized FACS to measure facial actions during conversations among stutterers and nonstutterers. As he predicted, the facial actions specified in FACS as relevant to anger occurred more often among the stutterers. There is little or no comparable evidence that the other facial measurement techniques listed in Table 2.3 can be used to measure spontaneous emotional expressions.

The only exception is Izard's use of his MAX technique to study infants. He found that observers scoring brief segments of videotape showing infant expressions *selected* to correspond to adult posed expressions could reliably identify the actions making up those expressions. This shows that his technique can be used to identify at least those particular expressions when they occur in spontaneous behavior. At this point, however, there is no evidence to support Izard's claim that an infant producing a particular expression is experiencing a particular emotion or blend of emotions. Because Izard has not described infants' facial behavior comprehensively, he cannot even specify how representative the selected expressions are in the behavior of infants of a given age and in a variety of situations.

Oster (1978; Oster & Ekman, 1978) has provided more complete information about the range of facial muscle activity observed in young

infants and about the young infant's capacity for coordinated facial movement. Unlike Izard, she began not by looking for adult posed expressions but by analyzing the configurations and sequences of facial actions actually produced by infants in a variety of situations. Oster found that almost all of the single facial actions included in FACS are apparent early in life. Though certain combinations of facial actions common in adult facial expression can be observed in the newborn period, others have not been observed in young infants. Oster (1978; Oster & Ekman, 1978) has argued that the only way to determine the affective meaning and signal function of infants' facial expressions is by a detailed description of the expressions themselves – including their timing and sequencing – combined with a thorough functional analysis of their behavioral correlates and stimulus context. Though far from complete, Oster's work has provided evidence that complex, spontaneous facial actions observed in young infants (e.g., smiling, brow knitting, pouting) are not random but represent organized patterns and sequences of facial muscle activity that are reliably related to other aspects of the infants' behavior (e.g., looking at or away from the care giver, motor quieting or restlessness, crying). Such relationships can provide insights into the infant's affective state and cognitive processes.

Conversational signals

Ekman and Friesen's FACS has been found useful in studying facial actions that play a role as conversational rather than as emotional signals.[10] Camras (1977) found differences in the syntactic form of questions that do and do not contain facial actions functioning as "question markers." Ekman, Camras, and Friesen (in preparation) found that the semantic context predicts which of two facial actions is used to provide speech emphasis. Baker (1979) used FACS to measure the facial actions shown by deaf persons when they sign. She has isolated particular combinations of facial actions that appear to serve syntactic functions.

Stable individual characteristics

Although Ermiane and Gergerian intended their facial measurement technique to differentiate personality and psychopathology, they have not reported any validity evidence. There is no evidence that any of the other facial measurement techniques are valid measures of any stable personal characteristic.

2.9. Costs

This last criterion for evaluating measurement techniques was not included in Table 2.3 because Ekman and Friesen was the only study to provide information about time costs for learning to measure and for scoring a specified sample of behavior. It takes approximately 100 hours to learn FACS. More than half of that time is spent scoring practice materials (still photographs and cinema) included in FACS at the end of each chapter in the instructional manual. Ekman and Friesen do not know whether people will still achieve high reliability if they skip such practice; they do know that high reliability was achieved when all the instructional steps were followed.

The costs for using a measurement technique once it is learned are much more difficult to estimate. For FACS and probably any other technique, the costs depend upon how densely the facial behaviors are packed in the time sample to be scored. Consider first comprehensive scoring in which FACS is used to measure *all* visible facial activity in a 15-second period. This could take as little as 1 minute if only one or two easily distinguished actions occurred and the investigator wanted only to locate start–stop points for each action. It could take as long as 10 hours, however, if the behavior was as densely packed as it is in the facial activity of deaf persons signing, and if onset–apex–offset was scored for every action. Ekman and Friesen have not observed any other instances in which facial behavior is so densely packed over so many seconds.

If selective rather than comprehensive scoring is done, the costs are lower. Presume that the investigator wants to score only actions that are said to be indicative of disgust, and he or she selects the actions listed in the *Investigator's Guide to FACS* (in Ekman & Friesen, 1978) that are predicted to be prototypic for that emotion. A 2:1 ratio, 30 seconds of scoring time for every 15 seconds of live action, is probably a reasonable estimate.

Ekman and Friesen have recently developed a more economical system for measuring the occurrence of single emotions, based on FACS. Occurrences of actions considered to be the most common signs of anger, fear, distress and/or sadness, disgust and/or comtempt, surprise, and happiness are noted. In what they call EMFACS (EM standing for emotion), time is saved in three ways:

1. Scoring does not extend to the particular action, but only to whether a member of a group of specified actions occurred. For example, there are seven signs grouped together that Ekman and Friesen consider relevant to disgust. EMFACS does not differentiate among nose wrinkling, nose

wrinkling plus upper lip raising plus lower lip depression, nose wrinkling plus lower lip elevation, and so on. If any of these is seen, a check is made for that grouping. All actions not in one of the groupings are ignored.
2. Intensity of action is not scored, although intensity is included in the requirements for particular actions within a grouping. For example, a slight depression of the lip corners with slight pushing up of the lower lip is included in the sad grouping, but when those two actions are moderate or strong they are not included.
3. The timing of actions is not measured; only a frequency count is taken. EMFACS takes one-fifth the time of FACS, but of course it suffers from all of the problems already discussed in detail for selective as compared to comprehensive measurement techniques.

Izard's MAX technique is similar to Ekman and Friesen's EMFACS. It, too, groups actions presumed to be relevant to the same emotion, and makes no provision for scoring the timing or the intensity of action. Unlike FACS, it requires the scorer to examine different regions of the face separately, and admittedly, it includes in some regions changes in appearance that are due to actions in another region. By contrast, FACS and EMFACS alert the scorer to all the appearance changes resulting from particular muscles. Rather than inspecting an arbitrary division of the face in three regions, the scorer learns where to look in the face for those changes. Izard's MAX technique was developed by collapsing some of the distinctions he had made in his earlier FMCS technique, but FMCS was itself selective, not comprehensive.

The virtue of EMFACS compared to Izard's MAX and other selective techniques is that what has been excluded is exactly specified. Work in progress by Ekman and Friesen will compare the validity of EMFACS with FACS scoring of the same videotapes of spontaneous facial actions obtained during interviews with depressed patients. This study will show how the two techniques compare in differentiating interviews at the time of admission to a mental hospital from interviews at time of discharge, in agreement with psychiatric diagnosis and in relationship to patients' self-reports of affect and mood.

2.10. Other techniques for facial measurement

EMG

A number of recent studies (see especially the work of Schwartz, Fair, Salt, Mandel, & Klerman, 1976a, 1976b) used surface EMG to measure facial activity in relation to emotion. In this procedure, quite small electrodes, about 1 cm in diameter, are taped onto the surface of the skin, which is first prepared by a slight scraping and application of paste or solution to enhance electrical contact. Wires or leads are run from the

electrodes to the recording machine. Four methodological difficulties are encountered in EMG measurement of facial activity.

First, the placement of leads on the face may itself inhibit facial activity. Movement of the head may loosen the electrodes, as may large facial muscular movements. To prevent these problems, subjects usually have been studied in isolation, or at least not when freely partaking in a conversation. Typically, subjects have been measured when trying to pose, imagine, remember, or create for themselves an emotional experience. Even in these situations, if a subject makes a large expression he will feel the tape that holds the electrode in place pull or tear. The use of surface EMG probably thus inhibits large expressions even if the experimenter does not explicitly do so by instruction or by choice of task for the subject to perform.

A second problem has to do with ambiguities about just what is being measured by surface EMG. Placing leads on the surface of the face often has the consequence, for most facial areas, of picking up activity in more than just the muscle targeted by the investigator. There is more than one muscle in most of these facial areas, and often their fibers interweave or they lie on top of each other. Although investigators using surface EMG have usually been careful to talk about a *region* rather than a muscle, their reasoning and much of their interpretation assumes success in isolating the activity of specific muscles. Ekman and Friesen, in a joint study with Schwartz (Ekman, Schwartz, & Friesen, in preparation), found that in the corrugator region the activity of many muscles other than corrugator itself was recorded by the electrode placed in this region: orbicularis oculi; levator labii superioris alaeque nasi; frontalis, pars medialis. The activity of these other muscles can be distinguished from that of corrugator, and they can be distinguished from each other, but these distinctions require more electrodes, some of which must be placed in adjacent facial regions. Another way to obtain measurement of specific muscles is to insert fine wires into a muscle, a procedure which, though not as painful as it sounds, is not practicable for many studies.

The third problem – whether EMG can provide measurement of more than just one or two emotional states – is fundamental to the complexity of facial activity. Most emotions cannot be identified by the activity of a single muscle. Happiness may be the only exception, but even here evidence (Ekman et al., 1980) suggests that the differentiation of felt from simulated happiness, of controlled from uncontrolled happiness, and of slight from extreme happiness requires measurement of more than one muscle. Disgust might be measured by the activity of two muscles, and surprise by the activity of three. To measure anger, or fear, or sadness, many muscles need to be measured. There are limits,

however, to the number of leads that can be placed on a person's face – limits dictated both by the necessity of monitoring so many channels of activity and by the number of wires that an investigator can paste on someone's face without being totally outrageous. The present state of surface EMG measurement is not likely to allow more than either the gross distinction between positive and negative affect or the targeting of only one or two emotions for study. (Just such findings have been reported for imagined and posed emotions.) Surface or even fine-wire EMG does not seem a method that lends itself to the study of situations in which an investigator wants to know about the occurrence of three, four, or more emotions, especially if the investigator does not wish to miss various manifestations of each emotion and blends among them. And, of course, EMG imposes the additional constraints of intrusiveness and limitations on the potential for movement.

Davidson (in press) raised a fourth problem common to studies using EMG to measure facial behavior. There is no standard system, as there is for EEG, for specifying exactly where to place an EMG electrode in order to detect activity in a particular facial region. Though investigators know roughly where each muscle is located, there is considerable latitude about exactly where to put an electrode. Without rather precise guidelines about electrode placement, research is vulnerable to error owing to unknown variations in electrode placement within and between subjects.

Consider the use of surface EMG to measure whether there is more or less activity in the zygomatic major region on the two sides of the face. Any differences obtained might not be due to the greater involvement of the right or left hemisphere but might to an unknown extent reflect differences in placement of the EMG electrode in relation to the muscle mass on the two sides of the face. Between-subjects designs, in which, for example, a measure of zygomatic major was correlated with a personality test score, would also be vulnerable to error owing to electrode placement. These problems can be circumvented by utilizing research designs in which EMG activity is compared in two or more conditions for each subject.

When EMG is used to measure change over time, and the leads must be placed on the face more than once, variations in placement of the leads on each occasion can introduce errors. Miller (1981–1982) has solved this problem by devising a template that can be attached to a subject repeatedly to ensure that electrode placement is identical on different occasions.

Surface EMG can play an important role in certain methodological studies of facial behavior. Mention was made earlier of Ekman and

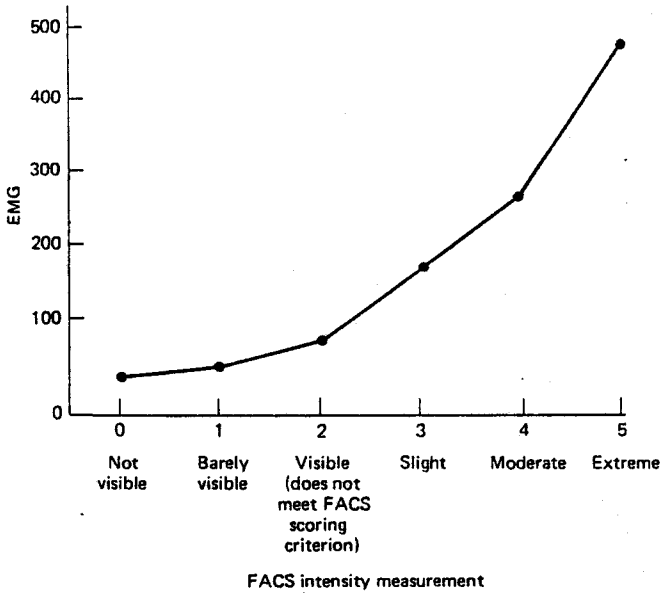

Figure 2.1. Plot of relationship between FACS and EMG measurement of performances of Action Unit 1 (frontalis, pars medialis)

Friesen's use of fine-wire EMG to stimulate and record facial movement in order to discover how the muscles work to change appearance. Surface EMG could be used to help teach people how the muscles work as part of the process of teaching them a visual measurement procedure such as FACS. Surface EMG can also be used to calibrate and investigate measurement of visible facial behavior. As mentioned earlier, Ekman and Friesen, in joint work with Schwartz (Ekman, Schwartz, and Friesen, in preparation), verified that the intensity scoring embodied in FACS is reliably related to changes in electrical activity. Persons highly skilled in the ability to activate specific muscles (Ekman and Oster) moved specific muscles on command at different intended intensity levels, while a video record was made and surface EMG was recorded. FACS scoring was later found to be highly correlated with the EMG readings (Pearson $r = 0.85$). Figure 2.1 shows an example from this data, a plot of the relationship between EMG measures of electrical activity and FACS scoring of the intensity of action for a specific muscle.

The major use for surface EMG is, however, not for such methodological studies, but for measures of phenomena that are difficult or impossible to measure with techniques based on visible movements. Ekman, Schwartz, and Friesen (in preparation) were able to show that there are reliable electrical changes associated with muscle tonus changes that are

not visible. For two muscles studied systematically (corrugator and frontalis, pars medialis), there were significant changes in EMG without any visible sign of activity when the performer was instructed just to think about each muscle. This study also showed that there are visible clues to muscle tension, measurable by EMG, when there is no movement. The persons measuring the faces with FACS guessed which muscle had been tensed when they could not see any movement. Sometimes the person guessing felt that there was no basis for the guess. At other times there seemed to be evidence of very slight tightening or bulging of skin. Analyses showed that when these guesses were correct – when the scorer predicted which muscle the performer was tensing, even though no movement was visible – there was a greater increase in EMG than when the guesses were incorrect.

EMG, then, may be the only method for measuring nonvisible changes in muscular tension, and for measuring changes that, while barely visible, involve not movement but bulging of the skin and would be hard to measure with any of the techniques described in Table 2.1.

Measurement of contour

Lasko (1979) has recently developed a method for measuring changes in the contour of different facial features (lips) or areas (infraorbital triangle). The researcher places a grid over each film or video frame in order to measure angles and area changes precisely. The method is designed to study changes resulting from muscular tension, blood flow gravity changes, or swellings owing to other causes. The technique appears promising for study of changes in appearance that are too small and too gradual to be measured readily with the techniques designed to measure movement. There is little reliability or validity data yet available, however. Also, the technique may be quite limited in its application, because only frames in which the subject's head position is exactly the same can be compared.

Other possible measures

There are other visible changes in the face that have not been systematically measured, that is, perspiration, blushing, and blanching. Thermal changes also could be measured, but no one yet has done so systematically.

2.11. Conclusions

This chapter has reviewed measurement techniques for only one type of signal: rapid, not slow or static. Among these, only one kind of rapid signal – visible movement – has been considered. Most of the studies that have used one or another technique to measure visible movement were concerned with only one of the many messages rapid signs may convey: information about emotion. Presumably, future research will expand to consider other messages and to develop methods for measuring rapid signals other than movement, as well as the variety of slow and static signals.

A few techniques have become available recently – those of Ekman and Friesen, Ermiane and Gergerian, and Izard. The first two were designed to be applicable to the study of any message, not just emotion. The availability of these techniques should encourage many more studies of facial movement. Wedding such studies of facial sign vehicles to studies using the more traditional message judgment approach should allow discovery of the particular actions that form the basis for correct and incorrect inferences when people judge facial expression. These techniques may also allow discovery of particular facial actions that are not customarily known or even knowable by the usual observer, movements that are too subtle and/or complex to notice or interpret when seen once at real time.

If research is generated by these facial measurement techniques, then the techniques themselves may well not survive: As a larger empirical base develops, it should become possible to improve, modify, or replace the techniques now in use. The methodological issues discussed in this chapter, however, should endure as guidelines for what to consider in developing or evaluating any procedure for measuring facial movement.

Appendix. How the facial action brow raise *is described in each of the 14 measurement techniques*

Birdwhistell
Raised Brows

Blurton Jones
A very conspicuous movement of raising the eyebrows which can be rather difficult to judge on photographs because of the individual variations in the resting position of the brows. One or more of the following criteria could apply:
a) The height of the brow above the eye corner appears to be equal or more than the width of the open eye. (Fig. 3a measure B equal or greater than A).

b) Horizontal lines visible across the forehead above the brows.
c) There is an enlarged area between the brow and the eyelids which is often highlighted (very pale) in photographs.
d) There is a less sharp fall from the brow into the eye socket (orbit) because the brow is raised beyond the edge of the orbit which it normally covers. Therefore there is less shadow between brow and eye than usual.
e) The shape of the eyebrows change, becoming more curved when they are raised (but they are not curved when the brows are slanted or oblique as well as raised).

Brow raising is presumably a result of contraction of the frontal belly of the occipto-frontalis, which can occur simultaneously with corrugator or orbicularis occuli contraction. Thus many oblique brows were also scored as raised.

Brannigan & Humphries

One or both eyebrows are raised and are held, at least briefly, in the raised position. They are not drawn in towards the midline and are not tilted.

Grant

The eyebrows are raised and stop in the raised position for an appreciable time (see plate 10A).

Flash. A quick raising and lower of eyebrows.

These two elements are very similar in use. They seem to have an attractive function, drawing the attention of the other person to the face. They are concerned with regulation and timing of speech.

Nystrom

– horizontal wrinkles
– elevated brows

(*Note*: These are listed by Nystrom as separate scoring items in his technique.)

Young & Decarie

Brow raise stare:

Brow: the eyebrows are raised and held giving them a curved appearance and creating horizontal creases on the brow. There is no inward movement of the eyebrows and no vertical furrow.

Eyes: The eyes may be held wide open but not sparkling, wrinkling at the corners and forming of pouching under the eyes. Blinking may be decelerated, and the head is definitely held in its regular forward position. Visual fixation on a specific target is characteristic of this expression.

Mouth: as in normal face.

Other: as in normal face.

(*Note*: Young & Decarie present this as a total face score. No provision is made for scoring if the brow raise action occurs without the eye action or with some other mouth action.)

Ekman, Friesen, & Tomkins

(*Note*: Two photographs depict this scoring item. The authors' Facial Affect Scoring Technique contains only visual, not verbal, descriptions.)

Izard: MAX (*Maximally Discriminative Facial Movement Coding System*)

Code 20: The brows are raised in their normal shape. The forehead shows some thickening and the tissue under the eyebrows some thinning out as a result of the eyebrows being raised. The thickening or massing of tissue in the forehead gives way to long transverse furrows with increasing age. The nasal root is narrowed. The skin directly below the eyebrows is stretched upward.

Code 30: The eyes have a widened and roundish appearance. The furrow above the eyelashes of the upper lid may be visible. The widened, roundish appearance of the eyes is brought about mainly by the eyebrow raise of code 20 that lifts and stretches the tissue between the eyebrow and the eyelid. The upper eyelid is not raised. The artist's drawing for 20 also illustrates 30.
(*Note*: Izard furnishes video examples of this action in addition to the artist's drawing.)

Ekman & Friesen: FACS (Facial Action Coding System)

There is one large muscle in the forehead area which raises the eyebrows. The medial (or central) portion of this muscle (Action Unit [AU] 1) can act separately from the lateral portion of this muscle (AU 2). The photograph on the left in Figure 2-1 shows the muscular basis of AU 1 and AU 2. The photograph on the right in Figure 2-1 shows the direction in which the muscle pulls skin when it contracts. The movement of AU 1 is to pull the medial part of the brow and center of the forehead upwards. The movement of AU 2 pulls the brow and the adjacent skin in the lateral portion of the forehead upwards towards the hairline. The combination of these two actions raises the inner (Action Unit 1) and the outer corners (Action Unit 2) of the eyebrows [2] producing changes in appearance which are the product of their joint action.

Appearance Changes Due 1 + 2

(1) Pulls the entire eyebrow (medial to lateral) upwards.
(2) Produces an arched appearance to the shape of the eyebrow.
(3) Bunches the skin in the forehead so that horizontal wrinkles appear across the entire forehead. The wrinkles may not appear in infants and children.
(4) Stretches the eye cover fold so that it is more apparent.
(5) In some people (those with deeply set eyes) the stretching of the eye cover fold reveals their upper eyelid, which usually is concealed by the eye cover fold.

Compare the photograph of AU 1 + 2 with the photograph of a neutral face.
Inspect the film depiction of AU 1 + 2

How to do 1 + 2

(*Note*: Ekman & Friesen's technique teaches learners how to perform each action so that they can utilize their own facial actions to understand the mechanics and appearance of the face.)

This should be easy to do. Simply lift your eyebrows up, both ends as high as you can. Note the wrinkling in your forehead. In some people the wrinkling does not occur but the skin is still bunched up. In some people these wrinkles are permanently etched (see photographs 0 and 0w) but they deepen noticeably when 1 + 2 acts.

Minimum requirements for scoring 1+2

The minimum requirements listed earlier in the MANUAL for scoring AU 1 alone and those for AU 2 alone are altered significantly in this combination.

(1) Entire brow raised *slightly*.

If you did not see the brows move it must also meet the additional requirements:

(2) *Slight* horizontal wrinkles or muscle bunching reaching across forehead. If horizontal wrinkles are evident in the neutral face, change from the neutral appearance must be *slight*.
and (3) *Slightly* more exposure of the eye cover fold than in neutral.
or (4) If there is no wrinkling or bunching in the brow, but the brow raise and exposure of the eye cover fold is *marked*, you can score 1 + 2.
(*Note*: The extent of action required by the terms *slightly*, *marked*, *extreme*, and

maximal is defined visually in both photographs and motion picture film examples.)

Frois-Wittmann
Brows raised.

Fulcher
Frontalis which raises the brows wrinkling the forehead transversely.

Ermiane & Gergerian
Frontalis – the eyebrow levator. Externalized emotionality.
(Raises the eyebrows).
Letting himself go to an impression.
(*Note*: A few photographic illustrations show this action.)

Landis
Frontalis. This is the vertical sheet muscle of the forehead, the contraction of which produces transverse wrinkles ("the wrinkled brow").

Acknowledgments

The preparation of this chapter was supported by a grant from the NIMH, MH 11976; a Research Scientist Development Award, MH 06092, from the NIMH; and a grant from the Harry F. Guggenheim Foundation. I am grateful to Wallace V. Friesen, Joseph C. Hager, and Maureen O'Sullivan for their many helpful suggestions and comments on earlier drafts.

Notes

1. Findings and hypotheses about the messages provided by static, slow, and rapid signals are discussed in Ekman, 1978.
2. Over the years I have proposed a number of different phrases to distinguish these two approaches. In previous discussions the *message judgment* approach has been labeled the stimulus, communicative, or judgment approach, and the *measurement of sign vehicles* approach has been labeled the response, indicative, or components approach. It is to be hoped that the present terms, taken from semiotics, allow a more lucid differentiation of these two methods.
3. Izard (personal communication, 1979) said that as part of an attempt to establish independent discovery, he deliberately did not examine Ekman & Friesen's Facial Action Coding System, even though it had already been published at the time when he was developing his measurement techniques.
4. See Kendon (Chapter 8) for a praiseworthy account of Birdwhistell, and Rosenfeld (Chapter 5) for a critique of Birdwhistell's methods.
5. Though neither of Izard's techniques (Affex or MAX) has been published as of late 1980, he has furnished information about both to those who inquire. Included are scoring manuals and illustrative material. An earlier version of MAX, FMCS (Facial Movement Coding System), is not available to others and is not discussed in this chapter.
6. Investigators studying the face of course do not agree about whether there is

definitive evidence regarding the particular facial actions that do and do not signal each emotion. Ekman and Oster (1979, p. 543), in reviewing the last decade of research on this topic, concluded that it is still a question of whether "facial expressions provide accurate information about the distinctions among several negative and positive emotions. The only evidence [indicating that] facial expressions [provide such discrete information] is for posed expressions." Izard (1979b) takes a much more positive view in describing what his measurement technique can do: "The content universe sampled by MAX consists of all the facial movements or appearance changes that signal affect" (p. 38). The evidence to support that claim is weak, however. One finding cited as evidence of validity is that scoring with MAX correlates with observers' judgments of emotion using Affex, but there is no validity evidence for Affex. The other validity evidence claimed is that infant facial expressions selected to correspond to adult poses and thus identified by MAX as representing one or another emotion are judged by observers to show those emotions. Here the limitation in evidence is threefold: (1) Judge agreement only establishes consensus; it does not demonstrate that the actions actually represent the emotions they are judged to show; (2) because the scoring technique was selective, there is no way of knowing whether other actions not scored might predict observers' message judgments just as well or better; (3) because the observers' choices were restricted, there is no way of knowing whether they would have described the infants' faces with the same emotion terms, or with any emotion terms at all, if they had been allowed free description.

7. In part because of its very uniformity, Ekman and Friesen consider the startle reaction to be not an emotion but instead a reflex. Other writers about emotion (e.g., Tomkins, 1962) disagree and classify startle with the emotion of surprise.
8. They acknowledge that for certain actions – for example, the movements of the tongue – their technique is not complete.
9. Frequency and duration measures also correlated with retrospective self-report, and the highest correlation was obtained with a score that combined intensity, frequency, and duration.
10. The distinction between emotional and conversational signals, with examples of how the same eyebrow movements can play either role, is given in Ekman, 1979.

References

Ancoli, S. *Psychophysiological response patterns of emotion.* Unpublished doctoral dissertation, University of California, San Francisco, 1979.

Ancoli, S., Kamiya, J., & Ekman, P. *Psychophysiological differentiation of positive and negative affect.* Paper presented at the annual meeting of the Biofeedback Association of America, Colorado Springs, 1980.

Baker, C. *Non-manual components of the Sign Language signal.* Paper presented at the NATO Advanced Study Institute on Sign Language and Cognition, Copenhagen, August 1979.

Birdwhistell, R. *Introduction to kinesics.* Washington, D.C.: Foreign Service Institute; Louisville: University of Louisville Press, 1952. Now available in microfilm only from University Microfilms, Inc., 313 N. First St., Ann Arbor Mich. Partly reprinted as an appendix to R. Birdwhistell, *Kinesics and context.* Philadelphia: University of Pennsylvania Press, 1970.

Blurton Jones, N. G. Criteria for use in describing facial expression in children. *Human Biology*, 1971, *41*, 365–413.
Brannigan, C. R., & Humphries, D. A. Human nonverbal behavior, a means of communication. In N. G. Blurton Jones (Ed.), *Ethological studies of child behavior*. Cambridge: Cambridge University Press 1972.
Camras, L. Facial expressions used by children in a conflict situation. *Child Development*, 1977, *48*, 1431–1435.
Davison, R. Psychophysiological studies of laterality. *Psychophysiology*, in press.
Duchenne, B. *Mechanisme de la physionomie humaine; ou, Analyse electrophysiologique de l'expression des passions*. Paris: Bailliere, 1862.
Ekman, P. Body position, facial expression and verbal behavior during interviews. *Journal of Abnormal and Social Psychology*, 1964, *68*(3), 295–301.
Ekman, P. Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology*, 1965, 2(5), 725–735.
Ekman, P. Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation* (Vol. 19). Lincoln: University of Nebraska Press, 1972.
Ekman, P. Facial signs. In T. Sebeok (Ed.), *Sight, sound and sense*. Bloomington: University of Indiana Press, 1978.
Ekman, P. About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology*. Cambridge: Cambridge University Press, 1979.
Ekman, P. Facial asymmetry. *Science*, 1980, *209*, 833–834.
Ekman, P., Camras, L., & Friesen, W. V. *Facial emphasis signals*. Manuscript in preparation.
Ekman, P. & Friesen, W. V. Nonverbal behavior in psychotherapy research. In J. Shlien (Ed.), *Research in Psychotherapy* (Vol. 3). Washington, D.C.: American Psychological Association, 1968.
Ekman, P. & Friesen, W. V. The repertoire of nonverbal behavior. *Semiotica*, 1969, *1*(1), 49–98.
Ekman, P., & Friesen, W. V. *Unmasking the face: A guide to recognizing emotions from facial clues*. New Jersey: Prentice-Hall, 1975.
Ekman, P. & Friesen, W. V. Measuring facial movement. *Journal of Environmental Psychology and Nonverbal Behavior*, 1976, *1*, 56–75.
Ekman, P., & Friesen, W. V. *The Facial Action Coding System: A technique for the measurement of facial movement*. Palo Alto, Calif.: Consulting Psychologists Press, 1978.
Ekman, P., Friesen, W. V., & Ancoli, S. Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 1980, *39*(6), 1125–1134.
Ekman, P., Friesen, W. V., & Ellsworth, P. *Emotion in the human face: Guidelines for research and an integration of findings*. New York: Pergamon Press, 1972 (2nd rev. ed.: P. Ekman, Ed., Cambridge University Press, in press).
Ekman, P., Friesen, W. V., & Simons, R. *Cognitive controls of the startle reaction*. Manuscript in preparation.
Ekman, P., Friesen, W. V., & Tomkins, S. S. Facial Affect Scoring Technique: A first validity study. *Semiotica*, 1971, *3*, 37–58.
Ekman, P., Hager, J. C., & Friesen, W. V. The symmetry of emotional and deliberate facial actions. *Psychophysiology*, 1981, *18*, 101–106.
Ekman, P., & Oster, H. Facial expressions of emotion. *Annual Review of Psychology*, 1979, *30*, 527–554.
Ekman, P., Schwartz, G., & Friesen, W. V. *Electrical and visible signs of facial action*. Manuscript in preparation.

Ermiane, R., & G
visage. Paris: I
Friesen, W. V., &
Frois-Wittmann, J
Psychology, 19
Fulcher, J. S. "Volt
of Psychology,
Grant, N. B. Hum
Hiatt, S., Campos
facial expressio
Research in C
Hjorstjo, C. H. M
Izard, C. *Facial e*
1979. (a)
Izard, C. *The ma*
Unpublished
Center, Univ
Izard, C. E., & D
facial express
emotions in in
1981.
Krause, R. Nonv
Zeitschrift fur
Landis, C. Studi
expression. *J*
Landis, C., & H
Winston, 193
Lasko, M. *A meth*
mathematical e
Lightoller, G. H.
rima oris wi
Anatomy, 192
Malmstrom, E., E
of surprise a
Psychologica
McGrew, W. C.
Press, 1972.
McGurk, H., &
26(4), 746–74
Mead, M. Reviev
Communicatic
Miller, A. J. Elec
Harvold (Ed
Cleft Palate Jo
Nystrom, M. Ne
and reliabili
1–16.
Oster, H. Facial
Rosenblum (
Oster, H., & Ek
(Ed.), *Minne*
Erlbaum, 19

Ermiane, R., & Gergerian, E. *Atlas of facial expressions; Album des expressions du visage*. Paris: La Pensee Universelle. 1978.

Friesen, W. V., & Ekman, P. (Eds.). *Analyzing facial action*. Book in preparation.

Frois-Wittmann, J. The judgment of facial expression. *Journal of Experimental Psychology*, 1930, *13*, 113–151.

Fulcher, J. S. "Voluntary" facial expression in blind and seeing children. *Archives of Psychology*, 1942, *38*(272), 1–49.

Grant, N. B. Human facial expression. *Man*, 1969, *4*, 525–536.

Hiatt, S., Campos, J., & Emde, R. *Fear, surprise, and happiness: The patterning of facial expression in infants*. Paper presented at the meeting of the Society for Research in Child Development, New Orleans, 1977.

Hjorstjo, C. H. *Man's face and mimic language*. Lund: Studentlitterature, 1970.

Izard, C. *Facial expression scoring manual*. [FMCS.] Unpublished manuscript, 1979. (a)

Izard, C. *The maximally descriminative facial movement coding system*. [MAX.] Unpublished manuscript, 1979. Available from Instructional Resources Center, University of Delaware, Newark. (b)

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

Krause, R. Nonverbales interaktives Verhalten von Stotterern. *Schweizrische Zeitschrift fur Psychologie und ihre Anwendungen*, 1978, *3*, 16–31.

Landis, C. Studies of emotional reactions: II. General behavior and facial expression. *Journal of Comparative Psychology*, 1924, *4*, 447–509.

Landis, C., & Hunt, W. A. *The startle pattern*. New York: Holt, Rinehart & Winston, 1939.

Lasko, M. *A method to measure curvature in facial expressional features: A complex mathematical description*. 1979. Unpublished manuscript.

Lightoller, G. H. S. Facial muscles: The modiolus and muscles surrounding the rima oris with some remarks about the panniculus adiposus. *Journal of Anatomy*, 1925, *60* (Pt. 1).

Malmstrom, E., Ekman, P., & Friesen, W. V. *Autonomic changes with facial displays of surprise and disgust*. Paper presented at the meeting of the Western Psychological Association, Portland, Oreg., 1972.

McGrew, W. C. *An ethological study of children's behavior*. New York: Academic Press, 1972.

McGurk, H., & MacDonald, J. Hearing lips and seeing voices. *Nature*, 1976, *26*(4), 746–748.

Mead, M. Review of *Darwin and facial expression*, ed. by P. Ekman. *Journal of Communication*, 1975, *25*(1), 209–213.

Miller, A. J. Electromyography in analysis of neuromuscular function. In E. Harvold (Ed.), *Hemifacial microsomia and other related craniofacial anomalies. Cleft Palate Journal Monograph*, 1981–1982, No. 1.

Nystrom, M. Neonatal facial-postural patterning during sleep: I. Description and reliability of observation. *Psychological Research Bulletin*, 1974, *14*(7), 1–16.

Oster, H. Facial expression and affect development. In M. Lewis & L. A. Rosenblum (Eds.), *Affect development*. New York: Plenum, 1978.

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

Oster, H., & Rosenstein, D. Measuring facial movement in infants. In W. V. Friesen & P. Ekman (Eds.), *Analyzing facial action*. Book in preparation.

Rubenstein, L. Facial expressions: An objective method in the quantitative evaluation of emotional change. *Behavior Research Methods and Instrumentation*, 1969, *1*, 305–306.

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. Facial expression and imagery in depression: An electromyographic study. *Psychosomatic Medicine*, 1976, *38*, 337–347. (a)

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. Facial muscle patterning to affective imagery in depressed and non-depressed subjects. *Science*, 1976, *192*(4238), 489–491. (b)

Seaford, H. W. *Maximizing replicability in describing facial behavior*. Paper presented at the meeting of the American Anthropological Association, Washington, D.C., 1976.

Tomkins, S. S. *Affect, imagery, consciousness*. Vol. 1: *The positive affects*. New York: Springer, 1962.

Young, G., & Decarie, T. G. An ethology-based catalogue of facial/vocal behaviors in infancy. *Animal Behavior*, 1977, *25*(1), 95–107.