Is There Universal Recognition of Emotion From Facial Expression? A Review of the Cross-Cultural Studies

James A. Russell

Emotions are universally recognized from facial expressions—or so it has been claimed. To support that claim, research has been carried out in various modern cultures and in cultures relatively isolated from Western influence. A review of the methods used in that research raises questions of its ecological, convergent, and internal validity. Forced-choice response format, within-subject design, preselected photographs of posed facial expressions, and other features of method are each problematic. When they are altered, less supportive or nonsupportive results occur. When they are combined, these method factors may help to shape the results. Facial expressions and emotion labels are probably associated, but the association may vary with culture and is loose enough to be consistent with various alternative accounts, 8 of which are discussed.

"Everyone knows that grief involves a gloomy and joy a cheerful countenance. . . . There are characteristic facial expressions which are observed to accompany anger, fear, erotic excitement, and all the other passions" (Aristotle, nd/1913, pp. 805, 808). Aristotle was not proposing a new idea but was cataloging what was known on the topic of physiognomy. The theory was that a person's physical appearance, especially in the face, reveals deeper characteristics: Poor proportions reveal a rogue, soft hair a coward, and a smile a happy person. [1]

Today, few psychologists share Aristotle's belief about the meaning of poor proportions or soft hair, but many share his beliefs about facial expression and emotion. Oatley and Jenkins (1992) observed, "By far the most extensive body of data in the field of human emotions is that on facial expressions of emotion" (p. 67). Recent reviews of those data (see Table 1) agree that the face reveals emotion in a way that is universally understood: Happiness, surprise, fear, anger, contempt, disgust, and sadness—these seven emotions, plus or minus two, are recognized from facial expressions by all human beings, regardless of their cultural background.

As Matsumoto (1990) said, "the universality of facial expressions of emotion is no longer debated in psychology" (p. 195).

Rather, it is a "fact," the implications of which are debated. Defending the idea of basic emotions against Ortony and Turner's (1990) criticisms, Ekman (1992b) and Izard (1992) pointed to universal facial expressions. Turner and Ortony (1992) replied, "We do not (and did not) dispute the fact that there are universal facial expressions associated with certain emotions" (p. 566).

If universality is a fact, then the implications are far-reaching indeed. D. E. Brown (1991) and Buss (1992) cited the existence of universal facial expressions as one of six key cases in their argument for bringing back the concept of human nature. For some theorists, universality is deeply revealing about the nature of emotion. For others, universality is a finding that must be accounted for by any viable theory of emotion (e.g., Oatley, 1992). Izard (1971, p. 188) wrote, "Emotion at one level is neuromuscular activity of the face" (emphasis Izard's). DePaulo (1992) wrote, "The fact that facial expressions of basic emotions are fundamentally the same across cultures (e.g., Ekman, 1972; Izard, 1971) is consistent with the position that there may be automatic links between the experiencing of the basic emotions and the expression of those emotions" (pp 205-206). Oatley and Jenkins (1992) pointed to facial expressions when they argued for the discreteness of emotions: "Investigators have found expressions specific to discrete emotions. Emotions may be considered discrete in the sense that they are produced and recognized pan-culturally" (p. 67). According to one proposal (although subsequently modified; Ekman, 1992a), whether a particular state is an emotion can be determined by that state's association with a universal facial expression: "There is a distinctive pancultural signal for each emotion. . . . If there is no distinctive universal facial expression associated with a given state, which functions as a signal, I propose that we not call that state an emotion" (Ekman, 1984, p. 330).

[1] Although the book quoted in this paragraph has been attributed to Aristotle, it may have been written or compiled by his students (Evans, 1969).

Table 1

Conclusions from Recent Reviews of Research on the Universality Thesis

Source	Conclusion
1. Ekman (1980)	"Are facial expressions of emotion the same for all human beings?" (p. 91). "Definitive data are now available on the question of universality" (p. 93). "There is conclusive scientific evidence to resolve the question of universality" (p. 96). "There are some facial expressions of emotion which are universal" (p. 137).
2. Izard (1980)	"Impressive evidence for the innateness and universality of six of the fundamental emotions: enjoyment (happiness), distress (sadness), anger, disgust, surprise, and fear" (p. 201). "Since all human beings recognize these expressions and attribute to them the *same* experiential significance, it is reasonable to infer that they are genetically based or preprogrammed" (p. 185).
3. Frijda (1986)	"Many facial expressions . . . occur throughout the world in every human race and culture. The expressions appear to represent, in every culture, the same emotions" (p. 67).
4. Fridlund, Ekman, and Oster (1987)	"Observers label certain facial expressions of emotion in the same way regardless of culture" (p. 157).
5. Gudykunst and Ting-Toomey (1988)	"The research cited indicates that facial expressions representing the basic emotions are recognized universally" (p. 182).
6. Buck (1988)	"Research on the communication of emotion via facial expression suggests that certain displays appear and are correctly recognized in widely different cultures" (p. 351).
7. Izard and Saxton (1988)	"The evidence for the innateness and universality of the expressions of the fundamental emotions is sufficiently robust to consider Darwin's hypothesis as an established axiom of behavioral science" (pp. 651–652).
8. Oster, Daily, and Goldenthal (1989)	"Conclusive evidence for the universality of certain facial expressions has come from studies in which observers were asked to identify the emotions shown in photographs of facial expressions" (p. 114).
9. D. E. Brown (1991)	"The conclusion seems inescapable: There are universal emotional expressions" (p. 26).
10. Mesquita and Frijda (1992)	"Certain facial expressions of emotion appear to be universal across cultures" (p. 14). "There appears to exist a universal human set of emotion reaction modes [including] facial expressions" (p. 21).
11. Carlson and Hatfield (1992)	"Ekman and other psychologists have uncovered compelling evidence that six basic emotions are expressed in much the same ways in all cultures" (p. 221).

The theories and ideas mentioned so far, their logical relations, and their evidentiary base are an important part of the psychology of emotion. The present article focuses on just one basic assumption: Emotions are universally recognized from facial expressions. After a brief historical sketch of and several comments on this thesis, I describe the key evidence on which it rests and raise questions about the methods used to gather that evidence. I then describe various alternative accounts of the data.

A Partial History

My view of the history of this topic is somewhat different than that usually presented. Most recent articles on facial expression begin not with Aristotle, but with Charles Darwin. The history of research on the universality thesis was described by Ekman (1980) in terms of Darwin's (1872/1965) ideas, challenges to Darwin, the rejection of Darwin, and, finally, "a conclusive answer to one of Darwin's questions, and an answer in agreement with Darwin's own conviction. There are some facial expressions of emotion which are universal" (p. 137). In his review,

Izard (1971) referred to the universality of facial expressions as "Darwin's hypothesis"(p. 225). Izard (1971) wrote that some researchers agreed that "Darwin was right" and others that "Darwin was wrong." Izard (1971) described one major review of the available evidence on facial expressions as "anti-Darwinian" (p. 225).

Darwin was not the originator of the universality thesis. Before Darwin's book appeared in 1872, facial expressions of emotion were a topic of great interest (Bain, 1855, 1859; Bell, 1806; Duchenne de Boulogne, 1862/1990; Piderit, 1867; Spencer, 1855) and were apparently assumed to be universal. For instance, in a book written before Charles Darwin was born, Bell (1806) wrote,

> The violent passions mark themselves so distinctly on the countenance, both of man, and of animals, that we are apt in the first instance to consider the movements by which they are indicated, as certain signs or characters provided by nature for the express purpose of intimating the internal emotion; and to suppose that they are interpreted by the observer in consequence of a peculiar and instinctive faculty. (Bell, 1806, p. 84)

Duchenne de Boulogne (1862/1990) referred explicitly to

universality. He attributed facial expression to a creator and wrote that

> once this language of facial expression had been created, it sufficed for Him to give all human beings the instinctive faculty of always expressing their sentiments by contracting the same muscles. This rendered the language universal and immutable. (p. 19)

In turn, Duchenne drew on even earlier work, such as that of the artist Le Brun (1702/1982). Nor was Le Brun the originator of the idea of universality. Aristotle's words quoted at the beginning of this article express an idea that can be found in Greek and Roman writers of antiquity (Evans, 1969). Darwin (1872/1965) drew on the work of his many predecessors (Montgomery, 1985) and stated himself that the universality of facial expression "has often been asserted," although "without much evidence" (p. 15). As a first step in his research program, Darwin thus sought evidence for the universality thesis "with all the races of mankind" (pp. 14–15). Still, Darwin's principal goal was not to establish universality, but to provide an evolutionary rather than creationist account of the origins of facial expressions. In short, the universality thesis predates Darwin by several thousand years. That facial expressions might not be universal appears not to have occurred to most people. Both evolutionists and creationists believed in universality. Universality is a background assumption, a part of common sense, at least in Western cultures. Whether the same holds true in other cultures is the topic of this article.

In the 20th century, the scientific study of emotion and facial expression was taken up by experimental psychologists. The earliest work (e.g., Feleky, 1914; Langfeld, 1918a, 1918b) did not cite Darwin but attributed to common sense the idea that emotions can be seen in the face. For example, Langfeld (1918a) commented, "It did not need this experiment to prove the well known fact that emotions and attitudes could be judged from pictures" (p. 183). Although some of the results agreed with this commonsense view, there soon appeared evidence that questioned whether emotions could be reliably judged from pictures of faces, even within our own culture. Different observers provided different emotion labels for the same expression (Buzby, 1924; Feleky, 1922). Observers were found to be swayed by the label that the experimenter suggested (Fernberger, 1928). The emotion attributed to a face could be altered by training (Allport, 1924; Guilford, 1929). When actual rather than simulated emotions were studied, the resulting facial expressions did not seem to reveal that emotion (Landis, 1924). The emotion attributed to infants depended more on the infant's situation than on facial behavior (Sherman, 1927). The 1920s were a low point in the fortunes of the universality thesis.

In 1938, a review and reanalysis of some of these data convinced Woodworth that disagreements over what emotion was shown in a particular face were more apparent than real. Earlier researchers had assumed that each emotion would have a unique facial expression. Woodworth first grouped together synonyms and words for closely related emotions (e.g., wonder, astonishment, amazement, and surprise). He then joined the groups into even broader clusters (e.g., joining the love, happiness, and mirth groups into one cluster, and joining the fear and

suffering groups into another). Scoring any response within the resulting broad cluster as correct eliminated much of the disagreement. When the six clusters were then ordered by similarity into a linear scale, most of the remaining disagreement was found to be minor. For example, disagreements in labeling the expressions of fear-and-suffering were not spread evenly across the five remaining clusters but tended to be only one step away: "Errors" were most often labels either from the surprise cluster on one side or from the anger and determination cluster on the other. When Feleky's (1922) seemingly negative judgment data were reanalyzed by Woodworth's method, the correlation between the emotion intended and that judged was .92.

Schlosberg (1941, 1952, 1954; Woodworth & Schlosberg, 1954) pursued Woodworth's approach, first altering the linear scale to a circular one. Schlosberg asked what the members of the broad clusters formed by Woodworth had in common. What, for example, unites love, happiness, and mirth? or fear and suffering? Schlosberg's suggestion was such things as degree of pleasantness versus unpleasantness. In his final version, Schlosberg (1954) suggested three such underlying dimensions: pleasant–unpleasant, degree of activation, and attention–rejection. Triandis and Lambert (1958) and Osgood (1966) took Schlosberg's idea cross-culturally and found similar dimensions.

The history of this topic has sometimes been presented as if it were a conflict between two opposing camps. For example, Ekman (1984) wrote,

> For more than 100 years scientists argued about whether facial expressions are universal or specific to each culture. On one side Darwin (1872/1965), and, more recently, Lorenz (1965) and Eibl-Eibesfeldt (1972), argued that facial expressions are innate, evolved behavior. On the other side, Birdwhistell (1970), Klineberg (1940), LaBarre (1947), Leach (1972), and Mead (1975), argued that facial expressions are instead like a language, socially learned, culturally controlled, and variable in meaning from one setting to another. (p. 319)

Various positions have actually been taken on these questions. Undoubtedly, writers of one extreme persuasion or the other can be found, but middle views can be found as well. As important examples of a middle view, I would cite Klineberg (1940), Woodworth and Schlosberg (1954), Bruner and Tagiuiri (1954), Triandis and Lambert (1958), and Osgood (1966).

First, consider Klineberg, the only psychologist cited in the quotation above and whom Ekman (1980) had earlier listed as one of Darwin's chief antagonists. Klineberg (1938) had reported that people in other cultures sometimes used unfamiliar facial expressions to signal their emotions, as when a Chinese novel described a character as sticking out his tongue in surprise. In his textbook, Klineberg (1940) detailed evidence of cultural variation, but he also wrote that "undoubtedly certain types of expressive behavior . . . are common to all human societies" (p. 176). He mentioned laughing with joy and tears in grief. Klineberg (1940) kept faith in the universality thesis in the face of cultural variation through three propositions. First, even if some facial expressions (such as the Chinese tongue protrusion) were culture specific, other expressions (such as laughing and crying) could still be universal; thus the culture specificity

of one expression did not speak against the universality of another. Second, culture might determine whether natural expressions were permitted, inhibited, or exaggerated. To illustrate, Klineberg (1940) cited large cultural variation in weeping from grief. This normative control later came to be called a *display rule*. Third, culture could determine what emotion occurred in a given situation. For example, having her husband take a second wife might be a source of pride in one society but might be a source of jealousy in another. Thus, witnessing their husband's second marriage, the wives in these two cultures would not show the same facial expression because they would not be experiencing the same emotion.

As further examples of a middle view, consider Woodworth and Schlosberg (1954) and their colleagues. In their concluding paragraph on facial expression, they disagreed with Klineberg's (1938, 1940) emphasis on cultural variation. They then wrote: "One emerges from a study of this topic with the conviction that there *are* certain basic emotional patterns in man" (p. 132, emphasis in the original). Triandis and Lambert (1958) and Osgood (1966) offered cross-cultural evidence for this conviction.

As a final example, consider Bruner and Tagiuiri's (1954) review, characterized by Izard (1971, p. 225) as "anti-Darwinian." Bruner and Tagiuiri (1954) pointed to the sobering negative findings that had accumulated, mainly in the 1920s, but they also pointed to positive findings:

> Other [researchers] have shown that emotional expressions were labeled with considerable accuracy (e.g., Darwin, 1872; Schulze, 1912; Feleky, 1914; Langfeld, 1918; Ruckmick, 1921; Stratton, 1921; Goodenough, 1931; Woodworth, 1938; Munn, 1940)." (p. 635)

Bruner and Tagiuiri (1954) emphasized Woodworth's (1938) and Schlosberg's (1941) analyses of these conflicting findings, apparently agreeing with the conclusion that "as a whole, judges of emotions from posed photographs do strikingly well" (Bruner & Tagiuri, 1954, p. 365). Their final assessment was not an extreme position:

> After considering the four technical problems, . . . we return finally to the question "To what extent are emotions recognizable?" We must come to the chastening conclusion that the literature is sufficiently haphazard to preclude a simple answer to this question. It depends. . . . That one can provide a multitude of situations in which accurate and consensual judgments can be obtained—of this there can be no question, if ever there was one. (p. 639)

In 1962, Tomkins published the first volume on his theory that was to mark the beginning of the modern era in the study of facial expression. Tomkins (1980) declared that "affect primarily is facial behavior." Tomkins (1962–1963) provided the list of so-called *basic* emotions that, with minor modifications, stands today and provided evidence that these emotions are easily read from facial expressions (Tomkins & McCarter, 1964). Tomkins inspired and tutored Ekman and Izard, who provided the most widely cited cross-cultural studies, including studies in isolated, illiterate societies (Ekman & Friesen, 1971; Ekman, Sorenson, & Friesen, 1969; Izard, 1971). Ekman and Izard also reviewed the relevant literature, describing their own vision of the history of the field (Ekman, Friesen, & Ellsworth, 1972;

Izard, 1971). For example, Ekman (1992b) wrote of the state of affairs in 1962: "Facial expressions were then considered to be an inaccurate source of culture-specific information" (p. 552). Ekman and Izard's ideas and cross-cultural studies entered introductory psychology texts in the late 1970s (Braun & Linder, 1979; Kagan & Havemann, 1976; London, 1978). In 1987, Ekman and his colleagues concluded that most psychologists had shifted to their position.

The Universality Thesis

Precisely what is meant by the *universality thesis?* Before any thesis along these lines can be verified, it has first to be formulated in a clear way. The universality thesis can be found asserted in textbooks, chapters, and journal articles, but it is not always clear exactly what is being asserted.

For example, precisely how did the conclusions regarding universality reached by Ekman (1972), Izard (1971), and their colleagues differ from the conclusions of Klineberg (1940), Bruner and Tagiuiri (1954), and Woodworth and Schlosberg (1954)? Ekman (1972) acknowledged the evidence of cultural variation and defended the universality thesis against such evidence through the same three ideas used by Klineberg (1940): Cultural variation could be accounted for (a) through some gestures being culture specific and others being universal; (b) through cultural norms that regulated when to mask, inhibit, or exaggerate natural facial expressions; and (c) through cultural influence on the cause of emotion. Although their predecessors' conclusions were cautious and complex and the new tone was more confident, the substantive difference was surprisingly difficult to state. It appeared to consist of two assertions: (a) Certain facial expressions are "easily recognized" (Izard, 1977, p. 501);[2] and (b) recognition is in terms of discrete, "specific emotion categories" (Izard, 1977, p. 502)—such as happiness, surprise, fear, anger, disgust, contempt, and sadness—rather than the overlapping broad clusters and dimensions suggested by Woodworth (1938) and Schlosberg (1941, 1952, 1954).

However, even these differences are not perfectly clear. What amount of agreement constitutes easy recognition? Under what circumstances is recognition predicted to be easy? And what is the precise difference between recognition in terms of discrete, specific emotion categories and recognition in terms of broad clusters? Izard (1971) counted a fair range of freely generated emotion labels as supporting his hypothesis. In contrast, Ekman and Friesen (1988) criticized Izard (1971) for this broad range and seemed to presuppose a one-to-one correspondence between a specific emotion term and a specific facial expression. Ekman (1992b) argued that the number of basic emotions could thus be revealed by studies of facial expression. Asking how to cut nature at the joints in the domain of emotion, Ekman (1992b) wrote, "I am tempted to say that those joints are na-

---

[2] Secondary sources have read this message clearly. For example, according to Carlson and Hatfield's (1992) textbook on the psychology of emotion, "sadness is a universal emotion easily recognized by all peoples, in all times, and at all places" (p. 266). "It is easy to recognize the face of anger. . . . The look is unmistakable" (p. 350).

kedly exposed by the findings on universality in expression" (p. 552). Yet, Ekman (1992a) also wrote not of specific discrete emotion categories, but of emotion families, thereby returning toward Woodworth's (1938) view.

Who said what on which occasion is less important than the question of what conclusion is warranted by the evidence. My focus in this article is not on the position of any one writer, but rather on how far the evidence goes. Let me propose therefore that the universality thesis be thought of as a family of hypotheses. Variations arise in three principal ways.

Specificity

First, the phrase *universality thesis* could describe propositions of varying specificity. At the nonspecific end of the continuum would fall the vague claim that some facial movements have some sort of emotional meaning, which is understood at above-chance levels by some humans in most cultures. At the other end of the continuum would fall the claim that, for example, exactly seven separate specific facial configurations signal the corresponding number of separate specific emotions, which are easily recognized as such by all humans. The conclusions reached by Klineberg (1940), Bruner and Tagiuiri (1954), and Woodworth and Schlosberg (1954) fall toward the nonspecific end of the continuum; conclusions by Izard (1971) and Ekman (1972) appear to fall toward the more specific end. More recently, Izard (1990) and Ekman (1992a, 1992b, 1993) may have taken somewhat different positions.

Ecological Relevance

The universality thesis could be taken to have various degrees of ecological relevance. A narrow interpretation would be that emotional facial signals can occur and can be recognized, but only under limited circumstances. A broader, more ecologically relevant, interpretation would be that unless there is a deliberate attempt at deception, the face naturally expresses the emotions that occur in everyday life and that observers routinely recognize the emotions expressed. In other words, this distinction is between competence and performance. Older evidence emphasized ecological relevance (Landis, 1924; Munn 1940; Sherman, 1927); more recent evidence, less so (Izard, 1971). Older, more cautious reviews were more concerned with ecological relevance (Bruner & Tagiuiri, 1954); more recent, more optimistic reviews, less so (Ekman et al., 1972; Izard, 1971). Izard (personal communication, November 10, 1992) emphasized a more narrow interpretation of his version of the universality thesis. In contrast, Ekman and Friesen (1975) emphasized the relevance of their findings to everyday situations, and Ekman, O'Sullivan, and Matsumoto (1991a) emphasized the question of ecological validity in evaluating research on facial expression.

Substance: Four Propositions

The universality thesis is ambiguous in yet another way. Four related propositions can be distinguished: (a) universality of facial movements: specific patterns of facial muscle movement oc-

cur in all human beings, (b) expressiveness of facial movements: certain facial patterns are manifestations of the same emotions in all human beings, (c) universality of attribution: observers everywhere attribute the same emotional meaning to those facial patterns (in a commonly seen phrase, those facial patterns have universal *signal value*),[3] and (d) correctness of attributions: observers are correct in the emotions they (consensually) attribute to those facial patterns. The last proposition, of course, presupposes the first three.

In much of the writing on this topic, these four propositions are not distinguished. Indeed, our everyday language invites their being confused. The phrase *facial expression* implies that facial movement has a meaning, which it expresses. The word *recognition* implies that what is recognized is really there. English words for facial movements, such as *frown*, are so imprecise that researchers are forced to speak instead of the "facial expression of fear," the "facial expression of anger," and so on. Thus, English words for facial movements imply their expression of specific emotions, which are there to be recognized or not.[4]

Less often, distinctions are noted, and an explicit argument is offered for combining these propositions. For example, Oster, Daily, and Goldenthal (1989) argued that

> if observers from very different cultures—including isolated, preliterate cultures—see the same emotions in the same faces, we can confidently conclude that these expressions have universal meaning. We can also infer that this meaning is derived from the observers' experience with spontaneous expressions observed in real-life, emotion-arousing situations in their own cultures. That is, we can infer that the expressions themselves are universal. (p. 114)[5]

---

[3] The word *attribution* is to be taken in as neutral a way as possible. There is no consensus on just what sort of process occurs when the observer "sees" an emotion in the face of another. Writers have used words such as *label, categorize, identify, see,* and *recognize*. The process might be innate or learned, conscious or unconscious, propositional or nonpropositional. On such questions, my review is noncommittal.

[4] To maintain some helpful distinctions, a convention is adopted in this article whereby emotion words used by observers are italicized (e.g., the word *happy*), but emotion labels used by researchers to refer to specific patterns of facial movement are in quotation marks (e.g., an expression of "happiness"). On the other hand, the words *recognition* and *facial expression* are used for lack of a simple alternative. *Recognition* simply means here a judgment that conforms to prediction. *Facial expression* simply means here facial movement.

[5] Caution rather than confidence is required in considering such arguments. Imagine that a man rubs his abdomen and, with both hands, rapidly pushes imaginary food into his mouth; that he cups his hands, throws back his head, and pretends to pour something down his throat; or that he closes his eyes and places the side of his head on his joined hands. It is possible that observers from very different cultures—including isolated, illiterate cultures—would see hunger, thirst, and sleep in these pantomimes. If that is so, then do the pantomimes have meaning derived from spontaneous actions observed in real-life situations and are the pantomimes themselves universal? The pantomimes presumably capture something of real-life action, but in an exaggerated, stylized form. One would not infer that the exact actions of the pantomime ordinarily accompany hunger, thirst, and sleep. Of course, facial expressions are not pantomimes, and my point here is a logical one. But this example does coincidentally underline a question that arises in the

Of course, these four propositions may be related empirically, and specific theories can predict links among them. But for analyzing the evidence, I believe that these four propositions are best kept separate, until proven otherwise. After all, it is at least conceivable that emotions are expressed by the face universally but that some cultures have failed to recognize this fact. Perhaps attribution of emotions to faces is a cultural achievement, which in our culture is partially correct but not perfect. Some other cultures might do better, and some worse. Alternatively, it is also at least conceivable that people falsely attribute emotions to facial movements. "Recognition" might be a widespread but false belief. Even when they are universal, folk beliefs need not be correct. It might not be wise for me to argue that any specific current folk belief is false, but it should be easy to see that widespread folk and even scientific beliefs of the past were false. Consider the belief, probably once universal, that the sun rises and sets on a stable earth or the belief that physiognomy reveals character. Consider the initial clash between common sense and scientific theories— such as the conception of space and time in modern physics, the age of the earth in geology, or the evolution of life in biology.

The Scope of This Article

A review of all the evidence potentially bearing on the universality thesis is beyond the scope of this article. I chose instead to focus on the evidence that has convinced past reviewers. Statements of the universality thesis found in journals, chapters, and textbooks are often followed by references to secondary sources. My review of this topic therefore began with secondary sources (Ekman, 1972, 1973; Ekman et al., 1972; Ekman & Oster, 1979; Izard, 1971; plus the more recent reviews listed in Table 1) from whom I extracted citations of primary evidence for universality. Their emphasis was clearly on modern cross-cultural judgment studies, which are therefore the topic of my review.[6]

By modern, I refer to studies beginning with Ekman et al. (1969). I excluded pre-1969 articles because Izard (1971) and Ekman et al. (1972) reviewed them in detail, often coming to conclusions different from those of previous reviewers and of the authors of the original articles. More recent reviews often ignored the earlier studies or cited the interpretations offered by Izard (1971) and Ekman et al. (1972). Further review of these disputed studies seemed less useful than a critical examination of more recent ones.

The principal limitation of my review is the focus on judgment studies. As implied by Oster et al.'s (1989) argument cited earlier, evidence that all "observers see the same emotions in the same faces" (i.e., the cross-cultural judgment studies, p. 114) has been taken as evidence for the entire universality thesis. Judgment data are offered to support the emotional meaning of facial behavior (e.g., Matsumoto & Ekman, 1988). In the following passage, Ekman (1980) captured the importance of the judgment study: "Definitive data are now available on the ques-

---

cross-cultural studies: Do the photographs of facial expressions typically used represent spontaneous, naturally occurring facial actions?

tion of universality" (p. 93). "The judgment study of the face . . . is the method employed in all but one of the experiments which have finally settled the issue of universality of facial expression" (p. 96).

I review studies of modern, literate cultures next and studies of isolated, illiterate cultures in a later section.

Judgment Studies in Literate Cultures

For literate cultures, the widely quoted high "recognition" scores came from eight studies. Different studies included slightly different lists of basic emotions, but all had six emotions in common. All results concerning these common six are listed in Table 2, separately for 20 Western (Europeans plus English, Spanish, or Portugese speakers in North or South America) and 11 non-Western (Asian and African) samples. "Recognition scores," which are the percentage of subjects who agreed with prediction, were greater than that to be expected by chance for each type of facial expression in each study. The expression of "happiness" achieved the highest averages, with a median of 96.4% in Western and 89.2% in non-Western cultures. The expression of "surprise" achieved the second highest, with a median of 87.5% in Western and 79.2% in non-Western cultures. Among the expressions for negative emotions ("sadness," "disgust," "anger," and "fear"), the range of medians was from 77.5% to 82.6% in Western and from 63.0% to 76.0% in non-Western cultures.

Three further studies have been cited but were not included in Table 2. First, Winkelmayer, Exline, Gottheil, and Paredes (1978) asked American, British, and Mexican observers to judge facial expressions of "happiness," "anger," and "sadness." Percentages of agreement were not reported for the emotions separately, however, and therefore were not reported in Table 2. The overall mean recognition score was 39.3%, whereas the figure to be expected by chance alone would have been 33%.

The second study (Ekman & Friesen, 1986) concerned a single emotion, contempt. Across 10 cultures, recognition averaged 75%. Results for individual cultures were not reported and therefore are not listed in Table 2. The basic result was subsequently replicated in six more cultures (Ekman & Heider, 1988; Matsumoto, 1992b). Although Ekman and Friesen's (1986)

---

[6] Various kinds of additional evidence were cited: naturalistic observation of facial movements in various cultures (Eibl-Eibesfeldt, 1972; Ekman, 1980); laboratory observation of facial movements of members of two different cultures (Ekman, 1972); correlation of facial movement with self-reports of emotion (Ekman, Friesen, & Anoli, 1980; but see Fridlund, 1991a); observation of facial communication in infants (Izard & Malatesta, 1987; but see Camras, 1992; Nelson, 1987; Oster, Hegley, & Nagel, 1992); observation of facial movements in children born blind (Eibl-Eibesfeldt, 1972; Goodenough, 1932; Thompson, 1941), and studies of the accuracy of judgments based on observation of facial movements (Ekman, 1982). Altogether, such studies suggested, to me, support for the first proposition listed above (that specific patterns of facial movement occur in all human beings), but not for the other three. For the purposes of this article, I assumed that the first proposition was essentially correct. This assumption left open the meaning of facial behavior.

Table 2
Recognition Scores Fom Eight Studies With Literate Subjects

		Facial expression					
Culture	*n*	"Happy"	"Surprise"	"Sadness"	"Fear"	"Disgust"	"Anger"
		Western cultures					
American[a]	99	97	91	73	88	82	69
Brazilian[a]	40	97	82	82	77	86	82
American[b]	89	96.8	90.5	74.0	76.0	83.2	89.2
English[b]	62	96.2	81.0	74.5	67.0	84.5	81.5
German[b]	158	98.2	85.5	67.2	84.0	73.0	83.2
Swedish[b]	41	96.5	81.0	71.5	88.8	88.0	82.2
French[b]	67	94.5	84.2	70.5	83.5	78.5	91.5
Swiss[b]	36	97.0	85.5	70.0	67.5	78.2	91.8
Greek[b]	50	93.5	80.2	54.5	67.8	87.5	80.0
Chilean[c]	119	90.2	88.3	90.9	78	85	76
Argentine[c]	168	94.0	93	87.6	68	79.3	71.6
Estonian[d]	70	88.0	82.5	84.7	60.2	89.0	77.7
American[e]	53	96.7	85.9	72.6	69.8	71.7	64.6
American[g]	40	100	92.5	87.5	67.5	92.5	90.0
Estonian[h]	85	90	94	86	91	71	67
German[h]	67	93	87	83	86	61	71
Greek[h]	61	93	91	80	74	77	77
Italian[h]	40	97	92	81	82	89	72
Scottish[h]	42	98	88	86	86	79	84
American[h]	30	95	92	92	84	86	81
Median		96.4	87.5	80.5	77.5	82.6	81.2
M		95.1	87.4	86.5	77.3	81.1	79.1
		Non-Western cultures					
Japanese[a]	29	87	87	74	71	82	63
Japanese[b]	60	93.8	79.2	66.8	58.2	55.8	56.8
African[b]	29	68.0	49.0	32.2	49.0	55.0	50.8
Kirghizian[d]	80	89.2	71.3	89.2	51.3	86.0	47.2
Malaysian[e]	30	95.8	69.8	66.4	45.6	59.2	49.8
Ethiopian[f]	100	86.8	50.5	52.0	58.8	54.8	37.3
Malaysian[g]	31	100	95	100	66.5	97.5	86
Chinese[h]	29	92	91	91	84	65	73
Japanese[h]	98	90	94	87	65	60	67
Sumatran[h]	36	69	78	91	70	70	70
Turkish[h]	64	87	90	76	76	74	79
Median		89.2	79.2	76.0	65.0	65.0	63.0
M		87.1	77.7	75.1	63.2	69.0	61.8

Note. Izard's (1971) term for "sadness" was *distress,* but it was defined as synonymous with *sadness.*
[a] Ekman, Sorenson, and Friesen (1969). [b] Izard (1971). [c] Ekman (1972). [d] Niit & Valsiner (1977).
[e] Boucher and Carlson (1980); figures given are unweighted average across two stimulus sets. [f] Ducci, Arcuri, W/Georgis, and Sineshaw (1982). [g] McAndrew (1986). [h] Ekman et al. (1987).

claim created controversy (Izard & Haynes, 1988; Ricci-Bitti, Brighetti, Garotti, & Boggi-Cavallo, 1989; Russell, 1991a, 1991c), Ekman and his colleagues defended their claim (Ekman & Friesen, 1988; Ekman et al., 1991a, 1991b; Matsumoto, 1992b).

The third study provided another Chinese sample. Chan (1985) showed nine of Izard's (1977) photographs of facial expressions to 124 medical students at the Chinese University of Hong Kong. The study was conducted both in Chinese and English. Subjects were allowed to select one or two words from the forced-choice list; the judgment was scored as correct if either choice was as predicted. These data were not included in Table 2 because this criterion for recognition was more lenient than

that used in the studies included there. Chan's obtained recognition scores were "happiness" 97.6, "surprise" 66.9, "sadness" 61.3, "anger" 96.0, "disgust" 62.1, and "fear" 66.9.

Some further judgment studies exist, but for various reasons they were not cited in the secondary sources as evidence for the universality thesis.[7] Some failed to report sufficient information

---

[7] Undoubtedly, unpublished data also exist. M. K. Mandal (personal communication, March 12, 1993) gathered unpublished data from 100 subjects in India, who were shown Ekman and Friesen's (1976) standardized set of photographs. The experiment was conducted in English. The results for the normative American sample and for the Indian sample, respectively, were 98.5 and 87.0 for "happiness," 92.3 and 75.1 for

to be useful. Some were criticized for their sample of facial expressions (Matsumoto, 1992a); for example, two (Cüceloglu, 1970; Kilbride & Yarczower, 1976) used line drawings of faces and are not reviewed here. Nevertheless, two sets of studies are useful.

First is additional data from Africa. Only two samples from Africa are shown in Table 2, and both suggest a noticeable discrepancy between African and Western results. Izard (1971) found an overall recognition score of 49.0% from a sample of Africans tested in France. This score contrasts with 83.4% in his normative, American group. He excluded the African data from most of his analyses (because the subjects were not tested in their native language). Ducci, Arcuri, W/Georgis, and Sineshaw (1982) obtained an overall recognition score of 51.9% from Ethiopians (no normative figure was reported, but the set of photographs was supplied by Ekman and Friesen). To estimate the reliability of this African–Western difference, it is helpful to look at additional data, which appear to replicate the difference even when the Africans were tested in their native language. Wolfgang and Cohen (1988) obtained an overall recognition score of 48% from Ethiopians (in contrast to 75% in their normative, Canadian sample). Kilbride and Yarczower (1980) obtained an overall recognition score of 62.5% from Zambians (in contrast to 94.2% in their normative, American group). Kilbride and Yarczower (1983) provided further data on Zambians. Although they did not report overall recognition scores and their set of facial expressions has been criticized (Matsumoto, 1992a), their results were informative. Given the same set of facial stimuli, which included both American and Zambian posers, Americans obtained significantly higher recognition scores than did Zambians.

Second is additional data from Japan. Izard (1971) and Ekman et al. (1987) had found a similarity between American and Japanese recognition for the hypothesized expressions of non-negative emotions, but a discrepancy on the "fear," "disgust," and "anger" expressions. Fortunately, three more studies have recently been published (Matsumoto, 1992a; Matsumoto & Ekman, 1988, 1989). Table 3 combines the earlier results, seen in Table 2, with the new ones. The additional data showed a similar pattern to the old. Moreover, Matsumoto (1992a) found a statistically significant interaction between culture and type of facial expression.

I now turn to a critical analysis of these studies. The results have already been summarized in Tables 2 and 3. The methods used are summarized in Table 4. The similarity of method across studies allows me to consider the studies as a whole rather than individually. For convenience, I call the combination of typical features the *standard method*. In the remainder of this section, I discuss first the results and then the individual features of the standard method. In doing so, I do not dispute the formal statistical finding in each study of an association between facial expression and emotion label. I do not seek one feature of method that is a fatal flaw. Rather, in the light of this discussion,

the evidence as a whole is later evaluated against the criteria of ecological, convergent, and internal validity. I focus on four questions: First, are recognition scores uniform across cultures? Second, are the results generalizable beyond the specific experimental context in which they were obtained? Third, is recognition in terms of specific emotions or broader clusters of emotions? And fourth, how easy is recognition, and is it possible that recognition scores were inflated by the method itself? One focus is thus the precise magnitude of the recognition scores. Evidence that agreement is lower than expected, even if still above chance, allows various alternative explanations. For example, agreement with prediction would be low but still greater than chance if observers could simply distinguish positive expressions from negative ones. Higher, but still less than perfect, agreement is predicted by Schlosberg's (1952, 1954; Woodworth & Schlosberg, 1954) model or its more modern variants (Russell & Bullock, 1987).

Results

Recognition scores from the eight studies of Table 2 were analyzed with a two-factor analysis of variance with a between–within design. Each sample was treated as a single case ($n = 20$ for Western cultures and $n = 11$ for non-Western cultures). Culture (Western or non-Western) was a between-subjects factor. Type of expression ("happy," "surprise," "sadness," "disgust," "anger," or "fear") was a within-subject factor. Adjustments for lack of sphericity were used where needed.

The main effect of culture was significant, $F(1, 29) = 15.36$, $p < .001$. The main effect of type of expression was significant, $F(5, 145) = 25.57$, $p < .001$. Their interaction was also significant, $F(5, 145) = 2.39$, $p < .05$. A simple effects analysis pursued the interaction effect obtained. When experimentwise alpha level was determined by Bonferroni's method, there were no significant effects of culture for "happy," "surprise," and "sadness" expressions, but there were for "disgust," $F(1, 104) = 9.17$, $p < .01$, "anger," $F(1, 104) = 18.94$, $p < .001$, and "fear," $F(1, 104) = 12.55$, $p < .001$, expressions.

A discriminant function analysis of the recognition scores showed significant, $\chi^2(1, N = 31) = 15.99$, $p < .001$, discrimination based on culture. Twenty-seven of the 31 groups were classified correctly. Four were misclassified. The Argentine and Estonian samples were misclassified with the non-Western set. (Estonian is a non-Indo-European language, and Estonia was a member of the Soviet Union at the time of the study.) The Turkish sample and McAndrew's (1986) Malaysian sample were misclassified with the Western set. Turkey straddles Asia and Europe and could have been included with the Western set. McAndrew's Malaysian sample were living in the United States.

Bonferroni matched-sample t tests with overall alpha set at .05 were used to explore the reliability of differences between recognition scores for the different types of expression. Among the Western samples, "happy" differed from each of the others; "surprise" differed from each of the others, except "disgust." Among the non-Western samples, "happy" differed from "disgust," "anger," and "fear."

A parallel set of analyses was conducted for the American–

"surprise," 88.6 and 73.0 for "sadness," 87.8 and 73.3 for "fear," 92.3 and 75.1 for "disgust," and 88.9 and 74.5 for "anger."

Table 3

Recognition Scores from Six Comparisons of Japanese and American Observers

Culture	n	"Happy"	"Surprise"	"Sadness"	"Fear"	"Disgust"	"Anger"
Ekman, Sorenson, and Friesen (1969)							
American	99	97	91	73	88	82	69
Japanese	29	87	87	74	71	82	63
Izard (1971)							
American	89	96.8	90.5	74.0	76.0	83.2	89.2
Japanese	60	93.8	79.2	66.8	58.2	55.8	56.8
Ekman et al. (1987)							
American	30	95	92	92	84	86	81
Japanese	98	90	94	87	65	60	67
Matsumoto and Ekman (1988)							
American	235	98.0	93.3	93.0	77.7	80.9	87.4
Japanese	154	98.0	89.2	75.6	37.6	70.1	68.1
Matsumoto and Ekman (1989)							
American	124	98.0	94.4	94.5	71.1	78.5	87.1
Japanese	110	97.6	88.0	77.1	30.8	68.2	69.6
Matsumoto (1992a)							
American	41	97.6	92.0	92.6	81.8	91.1	89.6
Japanese	44	98.3	92.0	71.9	54.6	74.7	64.2
Averages							
American							
Median		97.3	92.0	92.3	79.8	82.6	87.3
M		97.1	92.2	86.5	79.8	83.6	83.9
Japanese							
Median		95.7	88.6	74.8	56.4	69.2	65.6
M		94.1	88.2	75.4	52.9	68.5	64.8

Japanese comparisons shown in Table 3. In this analysis, the number of samples was even smaller ($n = 6$ within each culture), but the statistical results were similar. There were significant main and interactive effects in the main analysis: For culture, $F(1, 10) = 54.00$, $p < .001$; for type of expression, $F(5, 50) = 24.89$, $p < .001$; and for their interaction, $F(5, 145) = 4.50$, $p < .01$. The simple effects analysis did not yield significant effects of culture for the "happy," "surprise," or "sad" expressions, but did for "disgust," $F(1, 21) = 12.21$, $p < .01$; "anger," $F(1, 21) = 19.40$, $p < .001$; and "fear," $F(1, 21) = 38.49$, $p < .001$.

The discriminant function analysis was significant, $\chi^2(1, N = 12) = 12.00$, $p < .001$, and correctly classified all 12 samples on the basis of culture.

Bonferroni matched-sample t tests with overall alpha set at .05 were used to explore the reliability of differences among recognition scores for the different types of expression. Among the American samples, "happy" differed from each of the others except "sadness" and "anger." Among the Japanese samples, "happy" and "surprise" did not differ from each other, but each of them did differ from the rest. "Fear" also differed from "sad."

The analyses of the Japanese–American comparisons from Table 3 nicely complement the analyses of the more general comparisons of Table 2. Unlike the latter, the Japanese–American comparisons equate the number of samples, the stimuli, and other details of method. Altogether, the finding of reliable differences due to culture, to type of expression, and to their interaction appears to be robust. These analyses do not challenge the original researchers' conclusion that recognition scores are greater than chance, but they do show that recognition scores are not uniform. Recognition varies in a reliable and systematic fashion. The scores on the "happy expression" provide a baseline comparison of these observers' performance on the task per se. Therefore, reliable differences from this baseline in scores for other expressions in different cultures require examination.

Subjects

In all studies of literate cultures described in Tables 2 and 3, subjects were students: high school students in two samples and college students in the rest. Wolfgang and Cohen (1988) tested a sample of 96 volunteers from Central and South America living in Canada. The stimuli consisted of 40 posed facial expressions from a standardized set. Recognition scores varied with education level: Overall recognition was 81% from those with a university education, 66% from those with a high school education, and 43% from those with a primary school education. These differences were statistically significant.

There is also a question of how much the studies of Tables 2 and 3 provided a strict test of universality. The non-Western cultures included often had extensive contact with the West. The cross-cultural aspect of the studies is mitigated by the possibility that in these non-Western societies, students receive

Table 4
Methods Used in Studies With Literate Subjects

Source	Subjects	Facial stimuli	Design	Order of presentation	Response format
Ekman, Sorenson, and Friesen (1969)	College students	30 preselected still photos of posed and spontaneous expressions	Within subject	Unspecified	Forced choice, 6 alternatives
Izard (1971)	College students	36 preselected still photos of posed expressions	Within subject	1 random order, same for all Ss	Free label and forced choice, 9 alternatives
Ekman (1972)	Unspecified	Same as used in Ekman, Sorenson, and Friesen (1969)	Within subject	Unspecified	Forced choice, 6 alternatives, and intensity rating
Niit and Valsiner (1977, Exp. 1)	College students	35 preselected photos from Ekman and Friesen	Within subject	Unspecified	Forced choice, 7 alternatives
Winkelmayer, Exline, Gottheil, and Paredes (1978)	College students	60 Silent film clips, unselected spontaneous expressions	Within subject	1 semirandom order, same for all Ss	Forced choice, 3 alternatives
Boucher and Carlson (1980, Malay sample)	High school and college students	47 preselected still photos from Ekman and Friesen	Within subject	1 random order, same for all Ss	Free label and forced choice, 6 alternatives
Ducci, Arcuri, W/ Georgis, and Sineshaw (1982)	High school students	28 preselected still photos from Ekman and Friesen	Within subject	2 random orders, each counterbalanced	Forced choice, 7 alternatives
McAndrew (1986)	College students	30 preselected still photos from Ekman and Friesen	Within subject	1 random order, counterbalanced across Ss	Forced choice, 6 alternatives
Ekman and Friesen (1986)	College students	24 preselected still photographs of posed muscle configurations	Within subject	1 random order, same for all Ss	Forced choice, 7 alternatives
Ekman et al. (1987)	College students	18 preselected still photos of posed and spontaneous expressions	Within subject	1 random order, same for all Ss	Forced choice, 7 alternatives, plus quantitative rating

Note. Ss = subjects.

more exposure to Western culture than does the general public. To cite one of countless possibilities, students might be exposed to Western drama, and since the ancient Greeks, tragedy and comedy have been symbolized by frowning and smiling masks. Students might have seen Western books, magazines, newspapers, films, television, and other media; they might also have had more direct contact with Western students or faculty. The extent of this problem undoubtedly varied from study to study. Boucher and Carlson's (1980) Malaysian sample had no personal contact with Westerners, but McAndrew's (1986) Malaysian sample lived in the United States.

One study directly explored the effect of contact. Ducci et al. (1982) compared urban with rural high school students in Ethiopia. The urban students were in nearly continuous contact with Western culture, whereas the rural students were relatively isolated. Other than contact, most other relevant factors, such as age and education, were similar in the two groups. All subjects spoke Amharic and were equally familiar with the emotion terms used. Recognition scores were reliably greater for urban than for rural students.

For the more recent studies, the use of college students also raises an even more troubling possibility: The universality thesis or even the specific hypothesis proposed by Izard (1971) or Ekman (1972) could have been taught to some of the subjects who were used to test that thesis. Through their coursework or general reading, college students might have learned which specific

expressions had been hypothesized to be associated with which emotion. Books on drawing techniques describe how to draw a facial expression for each emotion (Faigin, 1990; Peck, 1987). Popularized books teach students how to recognize specific emotions from specific facial expressions (Ekman & Friesen, 1975). Darwin's (1872/1965) book might have been encountered in a psychology or biology class. The extent to which student subjects in each study were trained in the universality thesis is difficult to estimate. This problem would be greater if the subjects were drawn from psychology courses. For students enrolled in psychology courses up to the late 1970s, their textbook might have taught that situational information dominates over facial information in the judgment of emotion. After that time, some popular introductory psychology textbooks began to include material supporting the universality thesis, some with photographs of the hypothesized universal facial expressions (Braun & Linder, 1979; Kagan & Havemann, 1976; London, 1978). Table 2 shows no noticeable shift in recognition scores around the late 1970s, and only the Ducci et al. (1982) study directly supports these musings about training and cultural influence. Still, no evidence rules them out. In any future studies, the relevant background and training of the subjects should be examined and reported. As acknowledged by Izard (1977) and Ekman (1980), the question of cross-cultural universality is best answered by the study of cultures more isolated from Western influence than those of Tables 2 and 3.

Presentation of Stimuli

Psychological theory and research are unequivocal in proclaiming the dependency of a response to a given stimulus on other stimuli presented. This principle applies to responses as simple as recordings in single cells and to responses as complex as global judgments of happiness, morality, or beauty (Helson, 1964; Parducci, 1965; Russell & Lanius, 1984). In accord with this general principle, the judgment of a given facial expression has consistently been found to depend on what other faces are presented (Manis, 1967, 1971; Russell, 1991a; Russell & Fehr, 1987; Tanaka-Matsumi, Nelson, Attivissimo, & D'Urso, 1993; Thayer, 1980a, 1980b). Judgments of expressions claimed to be signals of basic emotions are no less relative: When particular stimulus contexts were created, the "contempt" expression was judged as disgust, the "surprise" expression was judged as a surprise–fear blend, and the "anger" expression was judged as sad. In addition, expressions claimed to be neutral have been judged to be happy when seen embedded in one stimulus set and as sad when seen embedded in another. The question then is how much the results summarized in Tables 2 and 3 depend on the particular stimulus context created by the experiment. Stimulus context might have exerted an influence through previewing, within-subject design, and order of presentation.

Previewing

In some studies (e.g., Winkelmayer et al., 1978), the subjects were first shown the entire set of facial expressions, after which each stimulus was redisplayed for the subjects to rate. In Izard's (1971) studies, 9 of 13 cultural samples performed the recognition task without previewing the photographs. Four of the groups were shown the expression photographs in a free-labeling task before the emotion recognition task. In Boucher and Carlson's (1980) study, observers saw the full set of expressions five times. Whether a preview of any kind was offered in other studies was not specified. Winkelmayer et al. (1978) reported that a preview was necessary for their subjects to achieve a reliable degree of agreement. Presenting subjects with the full range of facial expressions highlights both similarities within expression type and differences between expression types. Presenting an equal number of broad smiles, wrinkled noses, sneers, and so on further highlights similarities and differences.

Within-Subject Design

All studies in Tables 2 and 3 used a within-subject design: Each subject was asked to judge the entire set of facial expressions within a relatively short period of time. A within-subject design creates the same kind of problem as does previewing. It invites a more direct comparison between the various facial expressions than would be available in a between-subjects design (or in everyday encounters with facial expressions). Responding to a set of stimuli creates a series of complex assimilation and contrast effects.

Results from a within-subject design need to be verified with a between-subjects design. One attempt to do so failed: The fa-

cial expression published by Matsumoto and Ekman (1988) as a signal of contempt was judged as contempt in a within-subject design but was judged as disgust in a between-subjects design (Russell, 1991c). According to Ekman and Friesen (1986), a "contempt" expression is a unilateral raising and curling of the upper lip, and a "disgust" expression is a wrinkling of the nose. Seen separately (between-subjects design), these facial movements are both most typically labeled as disgust (Russell, 1991c). But when both expressions are presented to the same subject (within-subject design), then the subject might feel called on to notice the difference between the two expressions and (especially when the response scale calls for a distinction between disgust and contempt) to assign different labels to them. That they do so in a consensual manner is still informative, but the point is that method appears to influence the results obtained. Of course, contempt is the least well established of the "basic" emotions and, as noted, has aroused the most controversy. (In addition, Ekman et al., 1991a, argued that the ecological validity of my [Russell, 1991c] finding had yet to be established. I agree but believe that the finding nevertheless has methodological implications.)

Ekman et al. (1991b) recently revealed that "in our early studies (Ekman, 1972) we found unreliability in initial responses when subjects had to judge expressions. . . . We have found that subjects better understand what is expected of them after trying it a few times" (p. 294). Ekman et al. (1991b) reasoned that their subjects were initially unfamiliar with the task and that a few trials were required to understand the instructions. The high recognition scores for "happy" expressions in Table 2 speaks against this interpretation. Ekman's (1972) and Izard's (1971) theories would seem to predict that this would be an exceedingly simple task. (After all, the observer simply had to select one word from a short list for universal, biologically based signals.) In any case, if subjects must be trained on the task per se, then they should receive that training with neutral material (such as matching names to color patches), so that training does not influence the results of the experiment. The relevant point is that an initial response constitutes a between-subjects design and that this passage from Ekman et al. (1991b) therefore reveals that recognition scores were not as high (were not as reliable) in a between-subjects as in a within-subject design.

There are several more indications. Tanaka-Matsumi et al. (1993) obtained a recognition score of 44.2% for Ekman and Friesen's (1976) "sad" expression with a between-subjects design. Recognition of the "sad" expression rose to 58.3% when it was preceded by a single "happy" expression and to 72.1% when it was preceded by a single "anger" expression. They obtained a recognition score of 41.7% for Ekman and Friesen's (1976) "anger" expression with a between-subjects design. Recognition rose to 53.1%, 66.7%, and 57.9% when it was preceded by a single "happy," "sad," or "surprised" expression. Finally, I (Russell, 1993b) compared a between-subjects design with the combination of preview plus within-subject design. Although there was only one example of each expression and the dependent measure was free label rather than forced choice, the results (shown in Table 5) were consistent with the other evidence.

Table 5
Recognition Scores and Stimulus Presentation

Facial expression[a]	Design		χ^2 $(1, N = 224)$
	Preview + within S	Between S	
"Happy"	99.1	91.1	7.74**
"Surprise"	92.8	87.5	1.81
"Disgust"	79.5	66.1	5.07*
"Sad"	71.4	59.8	3.34
"Fear"	69.6	50.9	8.22**
"Anger"	60.7	40.2	9.45**
"Contempt"	1.8	0	
M	67.8	56.5	

Note. S = subject. Each recognition score is based on 112 observations. χ^2 was used to assess the effect of design, separately for each facial expression. The dependent measure was a free label. Data are described by Russell (1993). For "contempt," expected frequency was too low to calculate a chi-squared.
[a] Label is that provided by Matsumoto and Ekman (1988).
* $p < .05$. ** $p < .01$.

In all, what little evidence we have suggests that something about the within-subject design produces higher recognition scores than would a between-subjects design. I believe that the size of the effect varies with the type of expression. It was large for the "contempt" expression but was small for the six types of facial expressions included in Table 2 (given the other aspects of method). Nevertheless, the within-subject design, like previewing, may have increased the recognition scores there a bit.

Order of Presentation

For a within-subject design, order of stimulus presentation becomes an issue. Order should be varied systematically, or a separate random order should be created for each subject. However, the studies listed in Tables 2 and 3 sometimes used one or two orders of presentation. Different subjects were thus not responding simply to a particular facial expression alone. Rather, they were responding to that expression embedded within an identical context consisting of the preceding series of other facial expressions.

Giving different subjects the same order of presentation could inflate the amount of agreement among them. Inflation of the amount of agreement does not necessarily translate into increased agreement with prediction (indeed, it could do the opposite). Like the within-subject design itself, order of presentation alone challenges validity only insofar as increased agreement tends toward the hypothesis. Therefore, only in combination with other factors already sufficient to produce very high agreement with prediction could order of presentation exaggerate that agreement.

Conclusion

Exposure to the entire stimulus set, through previewing or a within-subject design, confronts the subjects with what might

appear to them to be a puzzle: how to assign, for example, seven labels to seven types of facial expression. If so, subjects might be guided in providing their responses or might even form a rough idea of the experimental hypothesis. Thus, stimulus presentation might account for some part (although not a large part) of the results generally offered in support of the universality thesis: Subjects give the same label to expressions of the same type and different labels to expressions of different types. Of course, if the facial expressions were completely meaningless to the subject, then the puzzle would remain unsolvable. But if a subject already had a fair idea of their meaning (such as appropriate labels for some of the expressions; or whether the expressions were positive or negative; or their values on Schlosberg's, 1954, dimensions; or labels for several expressions learned from the textbook), then previewing, within-subject design, and constant order of presentation could inflate the amount of agreement with prediction on some or all of the expressions shown.

Facial Stimuli

The facial expressions shown to subjects were unrepresentative of the population of facial expressions. In all studies of Tables 2 and 3, the stimulus expressions were preselected and, for the most part, posed. The researchers naturally had reasons for this approach, and responses to preselected posed expressions could address some questions. However, it is also important to ask about the possible consequences of preselection and posing. A full understanding of emotion and facial expression would include information about unselected spontaneous expressions.

Preselection

The exact degree of preselection is difficult to specify, but it was extreme. Ekman and Friesen (1971) described selecting their stimuli from over 3,000 photographs. Some of these 3,000, in turn, had been gathered by previous researchers, with each probably culled from still others. Izard's (1971) selection procedure was complex, beginning with a pool of about 1,000 posed expressions to which new ones were added. Photographs were eliminated from the pool if they failed to satisfy Izard himself or the original poser or if they failed to achieve 70% agreement with prediction from a panel of 25 to 30 judges. All but one of the remaining studies listed in Table 2 used photographs supplied by Ekman and Friesen, although not necessarily the same ones.

When the stimulus material is not so highly preselected, then recognition scores might not be as high.[8] Winkelmayer et al. (1978) used all of their stimuli and found considerably lower

---

[8] Izard (personal communication, November 10, 1992) reported that 34 of the photographs excluded by his procedure were subsequently shown to American and French subjects with a procedure similar but not identical to the standard method: "The mean percentage agreement for [these excluded] expressions was 31%. This contrasts with the agreement of 80% or better for American and French subjects viewing the photographs selected for the cross-cultural research." American and French observers agreed in their modal response for 17 of the 34.

114 JAMES A. RUSSELL

agreement with prediction than did any of the studies in Table 2, an average of 39.6% where chance would be 33.3%. Malatesta, Fiore, and Messina (1987) asked 14 elderly persons to pose facial expressions of happiness, fear, anger, sadness, and a neutral state. All 70 photographs (14 × 5) were then shown to 30 graduate students, who were given a forced-choice response scale. Overall, 41% of judgments were correct. This figure was influenced greatly by the happy poses, which were "almost always correct" (p. 65). With happy poses set aside, 28% of judgments were correct for the fear, anger, sadness, and neutral poses, where chance presumably would be 25%.

Posed Expressions

All studies cited in Table 2 used mainly or entirely posed rather than spontaneous expressions. The universality thesis is not, at least not directly, about posed faces. Posed faces do not express the emotion of the poser, but what the poser chooses to pretend and in a manner most likely to be understood by the observer. According to the notion of display rules, voluntarily posed expressions are culturally influenced and have been said to originate in a different region of the brain than do spontaneous facial expressions (Rinn, 1984). Posed expressions have been found to be more asymmetric than are spontaneous expressions (Skinner & Mullen, 1991). Reuter-Lorenz and Davidson (1981) found that subjects could tell when some of the expressions shown were posed. Posed expressions might be exaggerated or stylized. The poses for each type of expression might be more similar to each other, and more discriminable from poses of other types of expression, than spontaneous expressions. Posed expressions can be useful to answer some questions, but concerns of ecological validity raise the question of whether observers recognize the emotion conveyed by spontaneously produced facial expressions.

There is evidence that, at least in our culture, observers can accurately judge whether someone is feeling pleasant or unpleasant, or is in pleasant or unpleasant circumstances, on the basis of spontaneous facial expressions (Buck, Miller, & Caul, 1974; Buck, Savin, Miller, & Caul, 1972; Howell & Jorgenson, 1970; Nakamura, Buck, & Kenny, 1990). Can observers go beyond a distinction between pleasant and unpleasant? Three studies have addressed this question.

Ekman (1989). Ekman (1989) raised this question but cited only indirect evidence on the basis of a single, unpublished study. Emotion-inducing films were shown to 25 Japanese and 25 American viewers, whose faces were secretly videotaped. Specific facial movements such as smiling or lowering the brow were found on the videotapes in similar frequency in the two samples. Two segments from each videotape, one taken while the viewer watched a "neutral" travelogue and the other taken while the viewer watched a "stress" film (industrial accidents and the like), were then shown to a sample of judges. Judges were asked to guess whether the viewer was watching the travelogue or the stress film. According to Ekman (1972), "about 60% of the judgments" (p. 243) were correct (chance would have yielded 50%). Ekman (1972) argued that

these findings provide strong evidence in support of our position

that there are universal facial expressions of emotions. . . . It is reasonable to expect that the judges would make [their judgments] by a reasoning process in which they judged facial behavior as showing a particular emotion, for example, disgust, and disgust as being an emotion which would be more likely to occur during a stress film than a neutral film. (p. 244)

However, the small increase in accuracy might have been achieved without such reasoning. For example, judges might have interpreted facial expressions simply as positive or negative and guessed that negative expressions occurred during the stress film or positive expressions during the travelogue. Or the 60% accuracy might have been achieved with no reliance on facial expression at all. If the original viewers had, for example, averted their gaze more during the scenes of accidents than during the travelogue, then the judges might have exploited this difference. A test of Ekman's interpretation would require that judges agree on the specific emotional meaning of specific spontaneous facial expressions.

Motley and Camden (1988). Candid photographs of the facial expressions of 4 subjects were taken while they participated in six elaborately plotted interpersonal exchanges, each designed to elicit a different emotion. That the anticipated emotions were actually elicited was checked by retrospective report and physiological recordings. The same 4 subjects then posed facial expressions for the same six emotions (happiness, surprise, anger, disgust, sadness, and confusion). The resulting 48 photographs (6 spontaneous and 6 posed from each of 4 subjects) were then shown to 20 judges. For each photo, the judges were asked to select one emotion from a list of the six. For posed expressions, 81.4% of judgments were accurate, but for spontaneous expressions, 26.0% were. The difference was highly significant. Analysis of judgments of spontaneous expressions for individual emotions showed that judgments did not differ from random selection, except for happy expressions.

Wagner, MacDonald, and Manstead (1986). Wagner et al. (1986) videotaped the spontaneous facial expressions of persons viewing emotion-eliciting slides. The viewers indicated their response to each slide by selecting one of seven emotion words. Fifty-three judges then attempted to guess the viewers' emotions from the videotape by selecting one of the same seven emotion words. From their analysis of the results, Wagner et al. (1986) concluded that

overall accuracy was significantly greater than chance, although it was not impressive in absolute terms. Only happy, angry, and disgusted expressions were recognized at above-chance rates. (p. 737)

I believe that their modest conclusion was overly optimistic. Their three significant recognition scores were 48.4% for "happiness," 22.69% for "disgust," and 12.67% for "anger." As their level of comparison (what they called *chance*), Wagner et al. (1986) used the percentage of times a given label was used overall—that is, its base rate. Thus, 12.67% correct for "anger" was considered significant because it was reliably greater than 8.27%, the base rate with which the word *anger* was selected across all stimuli. Given the forced-choice response format and the nonrepresentative sample of stimuli, I find this comparison level necessary but insufficient. (Thus, judges might not so

Table 6

Responses to Negative Expressions From Wagner, MacDonald, and Manstead (1986)

Facial expression	Response word				
	Fear	Sad	Disgust	Anger	Total
Part 1. Percentage of all responses					
"Fear"	10.28	20.55	12.65	15.42	58.90
"Sad"	8.90	16.74	22.88	8.05	56.57
"Disgust"	5.26	18.69	22.69	9.98	56.62
"Anger"	9.67	18.33	19.00	12.67	59.67
M	8.53	18.58	19.31	11.53	
Part 2: Percentage of negative responses					
"Fear"	17.50	34.90	21.50	26.20	100.
"Sad"	15.70	29.60	40.40	14.20	100.
"Disgust"	9.30	33.00	40.10	17.60	100.
"Anger"	16.20	30.70	31.80	21.20	100.
M	14.68	32.05	33.45	19.80	

Note. Percentage of "correct" responses is underlined. From "Communication of Individual Emotions by Spontaneous Facial Expression" by H. L. Wagner, C. J. MacDonald, and A. S. R. Manstead, 1986, *Journal of Personality and Social Psychology, 50,* p. 741. Copyright 1986 by the American Psychological Association. Adapted by permission.

much have associated the label *anger* with facial expressions of "anger," as they did not find the word *anger* appropriate for smiles, for the wide open eyes of "surprise," and so on.) A more conservative approach would require judges to exceed not only base rate, but random choice. This is because observers could exceed base rate through a random process. For instance, imagine that for one type of facial expression, the judges had no idea whatsoever what emotion label to pick. If, on these trials only, judges in effect selected their responses blindfolded from a hat, then 14.3% of their responses (one out of seven) would be expected to be correct by chance alone. Return now to the three "significant" findings. Although happiness (48.4%) was well above this level of random choice, disgust (22.69%) was only somewhat above, and anger (12.67%) was slightly below. The obtained results cannot allow rejection of the possibility that subjects selected the words *anger* and *disgust* for the appropriate facial expressions as often as they did in a random process. (Wagner, 1993, subsequently proposed more sophisticated ways of assessing recognition.)

Even if performance exceeded both base rate and random choice, judges still might not associate specific facial expressions with specific emotions. Assume that the judges could tell positive from negative facial expressions reasonably well. This assumption is consistent with the markedly higher success on happiness both in this and in the Motley and Camden (1988) studies. In addition, other studies of spontaneous facial expressions show that judges can guess whether the eliciting circumstances were pleasant or unpleasant with accuracy greater than chance. The question then becomes whether judges discriminated among negative emotions—that is, beyond a positive-negative distinction.

Table 6 examines this question by focusing on those cases in which the viewer indicated a negative emotion and the judge rated the face as showing a negative emotion. Part 1 shows the figures reported by Wagner et al. (1986, Table 4). Part 2 converts those figures to percentages of negative responses. The simplest null hypothesis—that for these cases, choices among the four negative response alternatives were random—would predict that 25% of the responses in Part 2 would be correct. Overall, 27.1% of responses were correct. Percentages for "fear" and "anger" expressions fell below, "sad" and "disgust" expressions above, this level.

The seemingly better performance on "sad" (29.6%) and "disgust" (40.1%) expressions can now be compared with the base rates. Note that across the rows, the modal response was *sad* to the "fear" expressions and was *disgust* to the "sad," "disgust," and "anger" expressions. There was no differential labeling of "sad" and "disgust" facial expressions but a greater willingness to use the labels *sad* and *disgust*. The label *sad* was not more likely to be used for the "sad" expression than for any other negative expression—indeed, less likely. The label *disgust* was applied equally often to the "disgust" and "sad" expressions.

Conclusion. We have no cross-cultural studies of recognition of emotion from spontaneous facial expressions. Even in Western cultures, too few studies exist to draw firm conclusions. (Some earlier studies exist—Landis, 1924; Munn, 1940; Sherman, 1927—which were reviewed by Izard, 1971, and Ekman et al., 1972.) The available studies demonstrate the need to consider carefully the criterion by which performance is evaluated (Wagner, 1993). Apparently significant performance could be due to high base rate for specific labels, to random choice on certain trials, or to a positive-negative discrimination. (A further problem in these studies, to be discussed shortly, is their use of a forced-choice response format.) Beyond such factors, the available results (Motley & Camden, 1988; Wagner et al., 1986; Winkelmeyer et al., 1978) do not demonstrate a recognition of specific emotions from spontaneous facial expression.

Conversely, the available studies do not allow us to conclude that the fault—if *fault* is the right word—lies in the observer. In these studies, the person feeling the emotion might not have produced any facial expression in the first place. A study reported by Ekman, Friesen, and Ancoli (1980) suggests that at least some of the difficulty may indeed have arisen at the encoding stage. Something is amiss in spontaneous communication, but it could be either an encoding or a decoding problem. Perhaps further information on this question could come from scoring the available records of spontaneous expressions with one of the objective facial scoring systems (Ekman & Friesen, 1982; Izard, 1983).

Response Format

The results of Tables 2 and 3 were all derived from asking subjects to select one word from a prespecified list of emotions. For instance, a subject might have been asked to select one of the following: *happy, sad, afraid, angry, surprised, disgust,* or *contempt.* This type of forced-choice response format is a com-

mon instrument in research on emotion and has been advo-
cated as the method of choice in the study of facial expression,
although other response formats have also been used.

In this section, I first use a recent study to demonstrate that
problems can arise when forced-choice format is used—quite
different kinds of facial expressions can all be placed in the same
emotion category, and one and the same expression can achieve
consensus on quite different emotion categories, depending on
the choices given to subjects. Then I examine two alternative
methods of collecting subjects' responses: freely chosen labels
and quantitative ratings.

Forced Choice

A forced-choice format clearly alerts the subject to the exper-
imenter's expectation that the facial expression is to be inter-
preted in terms of emotion and even which emotions. After the
first trial, the set of category choices is primed and might influ-
ence subsequent perception (Wyer & Srull, 1981). Subjects were
not given the option of saying *none of the above* or of choosing a
nonemotion. For example, they were not allowed to give *frus-
tration, seeing a friend, attending to something novel,* or some
other situational interpretation to the face. They were not al-
lowed to describe the face as part of an instrumental response—
such as *threatening, staring,* or *avoiding.* They were not given
the option of choosing a more general emotional state, such as
unhappy, distressed, or *aroused.* Oster, Hegley, and Nagel
(1992) criticized judgment studies of babies' faces for lacking
just such nonspecific alternatives.

Given the list of options generally used, a forced-choice for-
mat is at odds with the properties of the emotion concepts pre-
sented (Russell, 1989b; Russell & Bullock, 1986). For instance,
forcing the observer to choose exactly one option treats the set
of options as mutually exclusive, which they are not: Subjects
place the same facial expression (or emotion of another or their
own emotion) into more than one emotion category. Forced
choice treats each option as an either–or (present–absent)
choice, which they are not: Subjects reliably rate different facial
expressions as belonging to a given emotion category to different
degrees.

Potential problems stemming from forced choice have not
been widely discussed in connection with the study of facial ex-
pression. If observers find on the prespecified list the precise
emotion label they have already spontaneously thought of, then
no serious problem arises. However, consider the situation in
which the list of response options fails to include a label for the
observer's spontaneous categorization. It might be thought that
such a situation would be immediately obvious because it
would result in idiosyncratic and thus random responses. Such
an outcome is possible, but—as will be shown—not inevitable.

A small study recently showed that forced choice could lead
to consensus on results rather different from those reported in
Table 2 (Russell, 1993a). The "wrong" expressions were cate-
gorized as a particular emotion when a different forced-choice
response format was used. By *wrong,* I mean not simply expres-
sions slightly different from the prototype. I mean expressions
previously claimed as expressing fundamentally different emo-

tions: the "sad" expression categorized as *fear,* the "fear" ex-
pression categorized as *surprise,* and so on.

There were 12 conditions—or demonstrations. In each con-
dition, subjects were shown one photograph of a facial expres-
sion from the Japanese and Caucasian Facial Expressions of
Emotion (JACFEE) set, developed by Matsumoto and Ekman
(1988). The subject was then asked to select one emotion label
from a list of options. (Slightly different methods were used for
the first 6 and last 6 conditions; five options versus six and the
placement of the predicted option.) In each condition, the pre-
dicted option was not a synonym for the emotion name sup-
plied by Matsumoto and Ekman (1988) but was predicted from
a circular model of the judgment space for emotions (Russell,
1980; Russell & Fehr, 1987). The options available in each con-
dition are shown in Table 7, in which the predicted option is
underlined. In the first condition, for example, the options were
happiness, surprise, contempt, fear, and *interest.* The subjects
were shown an "anger" expression and were predicted to select
contempt.

The percentage of subjects selecting the various options in
each condition is also shown in Table 7. Of course, some re-
sponse must be modal, but the results were not quantitatively
different from those seen in Table 2. The range of the 12 modal
responses was 46.3% to 96.3%, the median of which was 76.8%.
In each condition, responses were far from what would be ex-
pected by random assortment, and the predicted option was the
modal choice in each case.

In the first condition, subjects were shown an "anger" expres-
sion and were found to categorize it as *contempt*—at least 76%
did. This figure can be compared with previous results in which
the same type of expression was categorized as *anger.* For exam-
ple, Ekman and Friesen (1986) found that 73% of their subjects
categorized "anger" expressions as *anger.* In Table 2, the median
percentage of agreement for "anger" expressions was 81.2% for
Western subjects and 63.0% for non-Western subjects. Or the
present result can be compared with Ekman and Friesen's
(1986) claim that a very different expression, a unilateral lip
curl, is "unique to contempt" (p. 159): The percentage of their
subjects selecting *contempt* for the lip curl averaged 75% across
10 cultures.

In a subsequent condition, the "anger" expression was labeled
as *frustration* and, in another condition, as *disgust.* The "dis-
gust" expression was labeled *contempt* in one condition and *an-
ger* in another, the "sad" expression as *contempt* in one and as
fear in another, the "contempt" expression as *boredom* in one
and as *disgust* in another, the "fear" expression as *surprise,* and
the "surprise" expression as *fear.*

In no case did subjects select a category label at random, and
the use of the labels therefore implied no idiosyncratic selec-
tion. Forced choice could yield random selection, of course, but
the list of options used here was carefully constructed to allow
the modal response to be predicted. Even so, it should be clear
that these data do not show that the "anger" expression really
expresses contempt, that the "surprise" expression really ex-
presses fear, and so on. Nor do these results even say that ob-
servers necessarily attributed *contempt* to the "anger" and "dis-
gust" expressions, and so on. If that is so, then the same reason-

Table 7
Percentage of Judges Choosing Various Response Options

Facial expression[a]	Option		Option		Option		Option		Option		Option	
	Word	%	Word	%	Word	%	Word	%	Word	%	Word	%
Five choices												
"Anger"	Happiness	1.2	Surprise	6.3	Contempt	76.2	Fear	5.0	Interest	11.2		
"Disgust"	Happiness	0	Surprise	7.5	Contempt	90.0	Fear	0	Interest	2.5		
"Sadness"	Happiness	1.2	Surprise	1.2	Contempt	46.3	Fear	36.3	Interest	15.0		
"Anger"	Happiness	0	Surprise	0	Frustration	96.3	Fear	1.2	Interest	2.5		
"Contempt"	Happiness	2.5	Surprise	0	Boredom	88.8	Fear	0	Interest	8.8		
"Contempt"	Happiness	1.2	Surprise	1.2	Disgust	77.5	Fear	1.2	Interest	18.8		
Six choices												
"Anger"	Contempt	70.0	Joy	0.0	Relaxation	0.0	Surprise	5.0	Fear	10.0	Interest	15.0
"Anger"	Joy	0.0	Relaxation	0.0	Disgust	93.75	Surprise	0.0	Fear	0.0	Interest	6.25
"Fear"	Joy	0.0	Relaxation	0.0	Anger	5.0	Contempt	8.75	Surprise	71.25	Disgust	15
"Sadness"	Joy	5.0	Fear	73.75	Relaxation	2.5	Surprise	11.25	Excitement	0.0	Interest	7.5
"Disgust"	Joy	0.0	Relaxation	0.0	Surprise	12.5	Anger	75.0	Fear	7.5	Interest	5.0
"Surprise"	Joy	3.75	Relaxation	0.0	Anger	0.0	Contempt	6.25	Disgust	11.25	Fear	78.75

Note. Predicted response option is underlined. For each condition, a separate chi-squared test χ^2 ($N = 80$) examined whether response labels were chosen at random. For conditions with five choices, the degrees of freedom were 4. For conditions with six choices, the degrees of freedom were 5. For each condition, this null hypothesis was rejected at alpha of .001.

[a] Label is that given the photograph by Matsumoto and Ekman (1988).

ing shows that the results of Table 2 cannot be used to say that the various expressions really express the emotions claimed in those studies or that the subjects necessarily attributed those emotions to those expressions.

The method in this study is not advocated for general purposes. One anonymous reviewer commented that the response format used here was "as rigged" against the traditional hypothesis as the traditional format was "rigged in support." The purpose of this study was to expose potential problems. Given other options, subjects would have chosen other category labels. Given still other options, they would not be able to agree on one label as best. Thus, given different lists of options, forced choice can yield anything from random choice to a consensus, even on, from what researchers have generally concluded, the wrong answer.[9] Therefore, consensus achieved through forced choice need not point to subjects' unique response.

The question might be raised regarding whether the odd results here were due to the list of response options being overly restrictive. In general, any short list, including the ones used in the studies of Table 2, should be suspected of being overly restrictive. The results presented showed that the occurrence of near consensus was no evidence to the contrary. The question might be raised whether the odd results here were due to the list of response options failing to include the "correct" option. One problem is that we would need independent evidence to know the "correct" label, and such evidence is lacking for the studies of Table 2.

These results raise the question of whether observers would still prefer the "correct" (predicted) label when plausible alternatives were available. For example, shown Matsumoto and Ekman's (1988) "anger" expression, would observers select *anger* when the options were *anger, determination, frustration,*

hostility, hatred, jealousy, or *pain?* This question is relevant to whether observers interpret facial expressions in terms of specific emotion categories or broader clusters and was explored in a final study by Russell (1993a).

The study used two examples of Matsumoto and Ekman's (1988) "anger" expression and the seven response options listed above. These seven options had been chosen to be close to anger in Russell and Fehr's (1987) judgment space for emotions. Still, a semantic analysis leaves the word *anger* as the only option that unequivocally denotes the emotion of anger. *Hostility, hatred,* and *jealousy* refer to emotions qualitatively different from anger, according to Johnson-Laird and Oatley's (1989) semantic analysis. *Pain* refers to a complex sensation or perception rather than to a prototypical emotion. *Frustration* and *determination* refer to conditions of having a goal blocked and the resolve to do something. The results are shown in Table 8. The two photographs yielded somewhat different profiles—an important result itself because the photographs were predicted by Matsumoto and Ekman (1988) to signal the same emotion. Anger was not the modal choice for either expression.

In summary, forced choice can produce apparently contradictory results. For some facial expressions and some circumstances, subjects can consensually pick different emotion labels for the same expression. The universality thesis cannot be demonstrated solely through forced choice, and corroboration must

---

[9] Izard (personal communication, November 10, 1992) rightly pointed out that forced-choice format alone could not guarantee consensus. He pointed to unpublished data (see Footnote 8) on facial expressions that failed his selection criteria for inclusion in his set. Even with forced-choice response format, the mean percentage agreement on these photographs was 31%.

be obtained with other response formats. In the next section, I examine two alternatives: freely chosen labels and quantitative ratings.

Freely Chosen Labels

If recognition of emotion is easy and in terms of separate discrete emotion categories, then a list of emotions supplied by a researcher should not be necessary for recognition to occur. An observer's freely given response should reveal that recognition. However, when observers are given freedom to choose their own label, they don't always specify an emotion. This result was obtained by Darwin (1872/1965) when he asked over 20 persons to comment on still photographs. Darwin did not report the details of his results, but Frijda (1953) repeated this open-ended approach and found a variety of responses. The majority of responses did not mention emotions at all, but rather situations: "She looks as if she is looking at a small child playing." Another subject said that the person in the photograph appeared "as if looking at something with fixed attention, a game or something, tense, two cars which almost get into collision, but nothing happens. Then she says, 'Gosh, who would do anything so stupid!'" (Frijda, 1953, p. 314). Of course, these respondents might have used such situational descriptions to convey an emotional interpretation. But some might not have.

Subsequent "free label" studies have been less free. They required subjects to restrict their response to a word or two and told the subject that the word or two should describe an emotion. Of the five studies that obtained free labels, I discuss four here (Boucher & Carlson, 1980; Izard, 1971; Russell, 1991c; Russell, Suzuki, & Ishida, in press) and one more (Sorenson, 1975, 1976) in a later section.

Izard (1971). Izard (1971) obtained freely chosen labels from American, English, French, and Greek college students for 32 still photographs of eight different types of facial expression. These same students then provided the forced-choice judgments already cited in Table 2. (The degree of overlap in results from the two judgment tasks was not reported.) Recognition scores

Table 8
Percentage Selecting Different Labels for Two Facial Expressions Reported to be of Anger, From Russell (1993a)

| Response option | Expression | | |
	A	B	Total
Frustration	45.0	35.0	40.0
Determination	40.0	23.3	31.7
Anger	5.0	20.0	12.5
Hostility	6.7	11.7	9.2
Hatred	0.0	5.0	2.5
Jealousy	1.7	3.3	2.5
Pain	1.7	1.7	1.7

Note. $N = 120$. Sixty judges saw each photograph. From "Forced-Choice Response Format in the Study of Facial Expression" by J. A. Russell, 1993, *Motivation and Emotion, 17,* pp. 44 & 46. Copyright 1993 by Plenum Publishing Corporation. Adapted by permission.

Table 9
Recognition Scores From Forced Choice and Free Labels, From Izard (1971)

| Facial expression | American | | Greek | | No. words[a] correct |
	Forced	Free	Forced	Free	
"Interest"	84.5	36.9	66.0	35.8	22
"Joy"	96.8	89.8	93.5	80.2	28
"Surprise"	90.5	88.2	80.2	56.4	13
"Distress" (sadness)	74.0	60.1	54.5	49.6	28
"Disgust/contempt"	83.2	49.0	87.5	48.7	20
"Anger"	89.2	66.4	80.0	47.1	13
"Shame"	73.2	11.6	71.0	26.2	9
"Fear"	76.0	59.6	67.8	70.8	8
M	83.4	57.7	75.1	51.9	

Note. Figures given are percentages of responses considered correct. Sample sizes were 89 Americans and 50 Greeks.
[a] Number of words scored as correct in the free-label task; see Izard (1971) for the list. Izard (1971) reported separate results for men and women, which were combined here.

resulting from the two response formats are shown in Table 9 for two of the samples. Similar results were obtained from the remaining two samples.

Izard (1971) first told his subjects that the photographs were of people "who were trying to express an emotion. Some of the people tried to express a certain emotion; others tried to express another emotion, others still another, etc." (p. 268). Instructions then asked the subject to "decide which one emotion is expressed most strongly" (p. 268). Despite these instructions, some of the labels that subjects produced did not refer to an emotion (e.g., *deliberating, observation, clowning, smile, about to cry, bad news, skepticism,* and *sneer*). How many such responses occurred was not reported. Izard reported the responses I just listed because he considered them to be correct, and we do not know the nature of the responses (about 47% of the total) not counted as correct.

Ignoring words given by only 1 subject, Izard still found 224 different words or phrases produced (at least twice) for the eight types of expression. Izard therefore devised a scoring key in which, for example, 8 of the words were considered correct for fear, 28 for joy, and so on (the number correct for each emotion is given in Table 9). In all, 141 different words or phrases were scored as correct for the 8 emotions. Not all words that were considered correct for a given emotion were synonyms (e.g., *distress, loneliness, pain, pity,* and *worry* were considered correct for the "sad" expression; Izard's term for this expression was *distress,* but *distress* was considered a synonym for *sadness*). As noted, among the correct responses were also words not denoting emotions (*laugh, smile, clowning,* and *sees something pleasant* for "joy"; *bad news* and *crying* for "sadness"; *smirk* and *sneer* for "disgust/contempt"; *revenge* for "anger"; *repentance* for "shame"). The scoring method undoubtedly reflected the subjects' responses, but an issue that arose from the scoring was what version of the universality thesis the obtained results supported. It was not clear how these results supported Izard's

contention that recognition occurred in terms of "specific emotion categories" rather than Woodworth's (1938) broad clusters.

The results were not described fully. For example, the frequency with which specific labels were chosen was not reported, and it was therefore not clear whether the modal response for the proposed facial expression of "sadness/distress" was *distress, sadness, loneliness, worry* or what. Nonetheless, the results that were reported showed several clear trends. First, no marked differences appeared that were due to culture or language. Second, although a majority of free-label responses (about 53% overall) were considered correct, a large minority (from 42.3% of American responses to 48.1% of Greek responses) were considered incorrect. Third, recognition scores from free labeling were lower than from forced choice for every type of facial expression. Fourth, the difference between forced choice and free labeling varied with the type of expression. "Joy" replicated well; "shame" replicated poorly; and the other types of expression fell in between. There was a statistically significant main effect due to expression type (*emotion,* in Izard's, 1971, terms) for the correctness of free-label responses.

Boucher and Carlson (1980). Boucher and Carlson (1980) obtained both forced-choice (six options) and free-label responses from the same subjects, 30 Malays. Observers were shown facial expressions posed by Americans and Malays. With forced choice, the overall mean recognition score was 68.8% on the American and 60.0% on the Malay facial expressions. These figures were lower than had been obtained with the normative group of American subjects but were statistically significant for each emotion.

The free-label data were not described fully, but Boucher and Carlson (1980) did report that expressions for each emotion, except "fear," obtained recognition scores greater than chance and that free label had resulted in "an apparent lower level of accuracy" (p. 274) than had resulted from the forced choice. Thus, the recognition scores must have been lower than 60.0% to 68.8% but significantly greater than zero, except for the "fear" expression. It is therefore impossible to say how well the results supported the conclusion of high recognition. In this analysis, as in Izard's (1971), a response was considered correct if it was one of a set of words or phrases Boucher and Carlson considered similar to the label for the target emotion. Responses considered correct were not specified, and so it is impossible to say if the results more supported recognition in terms of specific emotions or broad clusters.

The degree to which the results from the free-label format replicated those from the forced-choice format varied with the type of expression. Boucher and Carlson (1980) asked how often the free label corresponded to the same subject's subsequently given forced choice to the same face (regardless of its correctness, i.e., of what emotion the poser intended). The freely chosen label was considered equivalent to the forced-choice response if it fell in the cluster of related terms. The proportion of forced-choice responses that coincided with freely chosen labels was as follows: *happy* 89.9%, *surprise* 60.7%, *anger* 60.4%, *sad* 47.7%, *disgust* 39.7%, and *fear* 7.7%.

Russell (1991c). In neither of the studies just reviewed were

Table 10

Frequency of Free-Response Labels for Expressions Reported to Be of Anger and Contempt

"Anger" expression		"Contempt" expression	
Label	Freq	Label	Freq
Frustration	49	*Disgust*	16
Anger	41	*Bored*	10
Mad	18	*Disappointment*	9
Constipated	4	*Puzzled*	6
Upset	3	*Confusion*	6
Confusion	2	*Frustration*	5
Making a decision	2	*Indifference*	5
Perturbed	2	*Smug*	5
Perplexed	2	*Contempt*	3
Irritable	2	*Perplexed*	3
Doubt	2	*Pissed off*	3
Pissed off	2	*Cynical*	3
Scorn	2	*Disgruntled*	3
Idiosyncratic	27	*Sarcastic*	3
		Anger	3
		Stupid	2
		Depression	2
		Indecisive	2
		Impatient	2
		Dissatisfaction	2
		Pain	2
		Troubled	2
		Arrogant	2
		Disbelief	2
		Perturbed	2
		Amused	2
		Idiosyncratic	46
		Other	9

Note. $n = 160$ for each type of expression.

the freely chosen labels reported in full. For example, it would be helpful to know the frequency with which different labels were chosen. Russell (1991c) reported freely chosen labels for the Matsumoto and Ekman (1988) facial expressions of "contempt." These data, along with comparable data for the Matsumoto and Ekman (1988) "anger" expressions, are summarized in Table 10. Each subject provided an emotion label of his or her own choice to one of 16 photographs (8 of "contempt" and 8 of "anger"). Twenty subjects responded to each photograph; total number of subjects was therefore 320. They responded to the question, "What mood or emotion is the woman [man] in the photograph feeling?" (p. 284).

Subjects were often reluctant to choose a single word; their initial responses were often a phrase or story, and the experimenter had to say, "If you had to choose a single word, what would that be?" (Even then, 9 subjects in the "contempt" condition claimed that they could not find a single word.) Single-word responses were tabulated. The result was 121 different responses, 40 for the "anger" expressions, 81 for the "contempt" expressions. (In this count, syntactic form was ignored.) The different labels that were produced by at least 2 subjects, and the frequency with which they were produced, are given in Table 10.

For the "anger" expression, the modal response was *frustration*, chosen by 30.6% of the subjects. The word *anger* was chosen by 25.6%. If *anger, mad, perturbed, irritable,* and *pissed off* are considered synonyms, then 40.6% gave the expected label. If *frustration* is added to the list of correct responses, then 71.2% of responses would be counted correct. On the other hand, including *frustration* raises a problem. Although frustration can lead to anger, it might also lead to sadness, shame, surprise, or some other emotion. (This point could be especially important in cross-cultural comparisons—for example, the Semai typically react to frustration with fear; Robarchek, 1977). Neither Tomkins and McCarter (1964) nor Izard (1971) included *frustration* as a correct response for their "anger" expression. The word *frustration* is not a synonym of *anger*. It is not clear that *frustration* refers to an emotion rather than to a situation. Oster et al. (1989) argued that frustration and anger have separate facial signals. The important point is not so much whether *frustration* is considered correct, but that a trade-off is implied by either course. If *frustration* is included, then more observers agree with the hypothesis, but the hypothesis has become less precise. If *frustration* is excluded, then the hypothesis being tested is more precise, but fewer observers agree with it.

For the "contempt" expression, the modal response was *disgust*, given by 10% of the sample. Only 2% of the sample gave *contempt*. More consensus would appear if terms were clustered. Any grouping of terms that would include responses from a majority of subjects would have to include terms that were more likely to refer to a different emotion (e.g., *disgust, anger,* or *sadness*) or a nonemotion (e.g., *puzzled* or *disbelief*) or to be so vague as to refer to almost any negative emotion (e.g., *troubled* or *dissatisfied*).

Russell et al (in press). Russell et al. (in press) obtained free emotion labels from 50 English-speaking Canadians, 38 Greeks, and 50 Japanese who were members of the general public rather than students. Observers were shown seven of Matsumoto and Ekman's (1988) photographs, reported to show universal facial expressions of basic emotions.

Responses again included nonemotions: *smile* for the "happiness" expression, *speechless* for "surprise," *disappointment* for "sadness," *frustration* for "anger," *having failed* for "disgust." To create a scoring system similar to that used by Izard (1971) and, presumably, Boucher and Carlson (1980), words reasonably close to the predicted term were scored as correct. The percentages of correct responses obtained are shown in Table 11. For comparison, although not a rigorous one, also given in Table 11 are Matsumoto and Ekman's (1988) normative figures from an American and from a Japanese sample for the same photographs. These normative figures were obtained with forced choice. For the "happy" and "surprise" expressions, free label replicated (indeed, surpassed) the forced-choice results. For the remaining expressions, forced-choice recognition scores were higher by from 9 to 45 percentage points in the English-speaking sample and by from 13 to 65 percentage points in the Japanese sample.

The English-speaking group gave the highest recognition scores based on free labeling to date: 67.7%, including 100% for the "happiness" expression and 96% for the "surprise" expres-

sion, indicating that subjects clearly understood the task. Overall agreement with prediction was 62.9% for the Greeks and 53.7% for the Japanese. (Because the "contempt" expression is controversial, it might be better to set aside the results for that expression. The free-label results from this study would then be higher, with median scores of 74%, 71.5%, and 68% in the Canadian, Greek, and Japanese samples.)

As in previous studies, recognition scores varied with type of expression. Overall, agreement was highest for the "happy," "surprise," and "sad" expressions. Agreement was low for the "contempt" expression, parallel to the Russell (1991c) result, and intermediate for "anger," "disgust," and "fear" expressions.

For four of the seven expressions, amount of agreement varied with the culture of the observer. Greeks produced similar scores to the Canadians, although the "surprise" expression was given a label from the *fear* cluster by 25% of the Greek respondents. The non-Western culture, Japan, yielded similar scores with two exceptions. The Japanese scores for "anger" (48%) and "fear" (14%) were significantly lower than in the Canadian normative group. The modal response to the "fear" expression was *odoroki* (surprise).

Conclusion. Four studies of freely generated labels yielded reasonably consistent results. Although recognition was generally lower than that obtained with forced choice, it was high enough for Izard (1971) and Boucher and Carlson (1980) to see the results as support for their hypotheses. In the Izard (1971) study, recognition was 53% overall, including that for happiness; for Boucher and Carlson (1980), the actual figure was not reported but was less than 60%–68%, again including happiness; recognition of the "fear" expression was no greater than chance. In Russell et al. (in press), recognition was 67.7% from Canadians, 62.9% from Greeks, and 53.7% from Japanese.

These average figures must be interpreted carefully. First, responses considered correct were not synonyms. When a certain percentage of subjects were reported to categorize a particular facial expression as, for example, *anger,* the more accurate statement would be that that percentage of observers used some word from a broad cluster related to *anger.* In their manner of scoring their data, Izard (1971) and Boucher and Carlson (1980) were defining the universality thesis in a way reminiscent of Woodworth (1938) and his followers (Schlosberg, 1952, 1954; Woodworth & Schlosberg, 1954). Scoring freely produced labels creates a trade-off: Higher recognition scores can be obtained at the cost of broadening the cluster of emotion words that count as correct. The experimenter can specify one particular emotion attributed to a given facial expression, but at the cost of reduced recognition scores. The amount of reduction appears to vary with the type of expression.

Second, recognition scores were not the same for every type of expression and culture. Recognition scores ranged from high for the "happy" expression to negligible for "shame" or "contempt." Scores for the expressions of "anger," "fear," and "disgust" were intermediate. Results were similar among Western samples (American, Canadian, English, French, and Greek) but lower in non-Western cultures (the Japanese and apparently

Table 11

Free-Label Judgments of Seven Facial Expressions, From Russell, Suzuki, and Ishida (in press)

| Predicted label of photograph[a] | English speaking | | | Greek | | Japanese | | |
| | Forced[b] | Free[c] | | Free | | Forced[d] | Free | |
	%	%	Mode	%	Mode	%	%	Mode
"Happy"	98.74	100	*Happy*	92	*Hara (joy)*	97.36	84*	*Tanoshii (delight)*
"Surprise"	95.34	96	*Surprise*	55*	*Ekplixi (surprise)*	90.26	94	*Odoroki (surprise)*
"Sadness"	94.48	70	*Sad*	75	*Lipi (sad)*	96.11	80	*Kanashimi (sorrow)*
"Anger"	91.92	78	*Angry*	63	*Thimos (anger)*	70.36	48*	*Ikari (anger)*
"Disgust"	75.25	66	*Disgust*	68	*Aidhia (disgust)*	74.45	56	*Iya (dislike, disgust, disagreeable)*
"Fear"	77.11	62	*Fear*	87	*Fovo (fear)*	37.66	14**	*Odoroki (surprise)*
"Contempt"	47.85	2	*Indifferent*	0	*Skeptikos (skeptical or thoughtful)*	64.94	0	*My hyoujou (expressionless)*
M		67.7		62.9			53.7	

Note. A chi-squared test was used to compare the free-label judgments from the Canadian sample with each of the other two, separately for each photograph.
[a] Label is that predicted by Matsumoto and Ekman (1988). [b] American normative sample, Matsumoto and Ekman (1988). [c] Canadian sample.
[d] Japanese normative sample, Matusumoto and Ekman (1988). From "Freely Produced Labels for Facial Expressions of Emotion" by J. A. Russell, N. Suzuki, and N. Ishida, in press, *Motivation and Emotion.* Copyright by Plenum Publishing Corporation. Reprinted by permission.
* $p < .05$. ** $p < .01$.

the Malaysians). For both Japanese and Malaysians, the "fear" expression was rarely labeled as *fear.*

Lower recognition scores with free-label format have been attributed (Izard, 1971) to the difficulty of free labeling (a production measure) rather than to problems with forced choice (a recognition measure). But Izard (1971) provided no comparison task that would allow an assessment of the difficulty of free label.[10]

Moreover, free label is not without its own problems. The instructions and procedure emphasized that the label should describe an emotion. Free label, like forced choice, presupposes that the number of labels needed is one and that emotion labels apply to facial expressions in a dichotomous fashion. That is, judges are forced to choose one label (albeit from an unrestricted set) as applying completely and thereby implying that all other labels do not apply. These restrictions imposed on observers happen to correspond to the assumptions of certain versions of the universality thesis. Relaxing these constraints would allow a judge to say, for a particular face, that more than one label applies and that a given label applies only to a certain degree.

Quantitative Ratings

Given the problems just mentioned, quantitative ratings on multiple scales are a needed complement to the forced-choice and free-label formats. Quantitative ratings would not be biased against any specific hypothesis but would simply be more neutral. If a specific facial expression is a highly recognizable, pancultural signal unique to a specific emotion, then subjects could use quantitative scales to judge it as expressing that emotion to a high degree and as expressing all other emotions to a low degree.

Ekman et al. (1987) acknowledged that problems might exist with the standard forced-choice format and therefore gathered additional data from each subject. They hypothesized that similar results would occur "even when observers were allowed to choose more than one emotion" (p. 714). Each observer was first given the standard forced choice (among seven alternatives) and then asked to make quantitative ratings on any of the seven that were present in the face. Ekman et al. (1987) analyzed these judgments by considering only the highest rating given on each trial. They reported the following:

> We determined whether the emotion with the most intense rating was the emotion predicted by Ekman and Friesen and was the same across cultures. [The hypothesis] was supported; in 177 of 180 times, the emotion rated strongest by the largest number of observers in each culture was the predicted emotion. This is the first evidence of cross-cultural agreement about the most intense emotion when observers can choose more than one emotion. (p. 715)

No further details were reported for these data. Ekman et al. (1987) did not report how they scored ties (cases in which a subject gave the same quantitative rating to two emotion terms), and ties are frequent when subjects give quantitative ratings. As

[10] In unpublished data (Russell, 1993b), I sought some baseline with which free labels for facial expressions could be compared. I asked subjects to label photographs of three types: facial expressions drawn from Matsumoto and Ekman (1988), vehicles, and animals. The recognition scores were as follows: for facial expressions, happy 91.1%, surprise 87.5%, fear 50.9%, sad 59.8%, anger 40.2%, disgust 66.1%, and contempt 0%; for vehicles, motorcycle 100%, truck 98%, boat 100%, bus 100%, train 96%, airplane 100%, automobile 80%; for animals, elephant 100%, duck 89%, horse 100%, cow 100%, dog 100%, chicken 96%, and pig 100%.

Table 12

Mean Rating on Accuracy of Category as Descriptor of a Facial Expression

Category label	Facial expression					
	"Interest"	"Sadness"	"Disgust"	"Anger"	"Fear"	"Surprise"
Excited	4.5	3.0	4.3	3.4	3.5	3.6
Happy	5.1	2.1	2.3	1.7	2.2	3.8
Calm	5.2	4.0	3.1	4.4	3.6	5.0
Sleepy	2.5	3.1	2.2	3.0	3.3	2.4
Bored	3.2	3.3	3.1	4.4	3.4	3.5
Sad	3.1	5.9	3.2	5.3	4.5	4.0
Disgusted	3.1	4.6	6.8	5.6	4.4	3.0
Angry	2.9	4.8	5.7	6.0	5.2	2.9
Afraid	3.3	5.8	3.7	4.9	5.8	4.3
Surprised	4.8	4.2	3.6	3.0	4.7	6.6
Reliability	.89	.95	.96	.95	.93	.85

Note. Ratings were made on a scale from *extremely inaccurate* (1) to *extremely accurate* (8). Label was that provided by Ekman. The plate number in Ekman's (1980) book for each facial expression is as follows: "interest" 22, "sadness" 14, "disgust" 16, "anger" 17, "fear" 10, and "surprise" 7.

analyzed, these data do not go far beyond the forced-choice data. Just before giving the quantitative ratings for each face, each college student subject was asked to choose one of the emotion terms as best. It is not surprising that the same subject would then give his or her highest quantitative rating to that same emotion term.

Analysis of other quantitative ratings shows that subjects do not interpret facial expressions dichotomously. Rather than rating one emotion high and all others low, subjects report that labels vary gradually from the best descriptor to the worst. Russell and Bullock (1986, Study 4) asked 50 subjects to provide quantitative ratings for six photographs taken from Ekman (1980). The results are shown in Table 12. The highest mean ratings were as Ekman (1980) predicted, but subjects did not give low ratings to all other labels. One example is provided by the column for Photo 17, presented by Ekman as "anger." On a 1–8 scale, this face was rated 6.0 for *anger*, but 5.6 for *disgust*, 5.3 for *sad*, 4.9 for *afraid*, and so on in a reliable order. Similarly, the "sadness" expression was rated 5.9 on *sad*, but 5.8 on *afraid*; the "disgust" expression was rated 6.8 on *disgust*, but 5.7 on *angry*; the "fear" expression was rated 5.8 on *afraid*, but 5.2 on *angry*.

Russell (1991c) reported quantitative ratings on six emotion scales for the Matsumoto and Ekman (1988) "contempt" expressions. Mean ratings, along with comparable results for the Matsumoto and Ekman (1988) "anger" expressions, are shown in Table 13. For the "contempt" expressions, ratings on *contempt* were not high, either in an absolute or in a relative sense. Absolute ratings on a 1–4 scale for the contempt label were low: The mean, 2.2, was just above *slightly* and below *moderately*. The *contempt* ratings were also low in relation to some others. Although *contempt* ratings were higher than *anger* ratings, they were not higher than ratings on the remaining labels and, indeed, were lower than on *disgust* and *boredom*.

For the "anger" expressions, the mean rating on *anger*, 2.8, was also between "slightly" and "moderately." Although *anger*

ratings were higher than on most of the other labels, they were lower than on *frustration*. The result for *frustration* replicates the free-label result of Table 10.

Conclusions on Response Format

The methods used so far to collect observers' judgments are highly reactive, and each has problems. Little is known about the observer's spontaneous response. Knowing more about the

Table 13

Quantitative Ratings of Facial Expressions Reported to Be of Contempt and Anger

Label	"Contempt" expressions			"Anger" expressions		
	M	SD	t^a	M	SD	t^b
Boredom	2.75	1.14	3.04*	1.17	0.52	11.87**
Disgust	2.58	0.92	3.01*	2.19	0.81	3.83**
Frustration	2.30	0.90	0.79	3.06	0.92	1.40
Contempt	2.17	0.79	—	2.17	0.92	4.97**
Scorn	2.05	0.98	0.96	2.44	0.97	2.72*
Anger	1.64	0.84	4.15**	2.83	0.97	—

Note. $n = 64$ for each type of expression. The stimuli were 3 × 5 color prints of photographs from the set Japanese and Caucasian Facial Expressions of Emotion (JACFEE) developed by Matsumoto and Ekman (1988); eight were of "contempt," and eight were of "anger." The response scale was labeled *not at all* (1), *slightly* (2), *moderately* (3), and *extremely* (4). A repeated measures analysis of variance (ANOVA) showed that subjects responded differently to the different labels, $F(5, 315) = 12.09$, $p < .001$, for the "contempt" expressions and $F(5, 315) = 39.55$, $p < .001$, for the "anger" expressions.
[a] t test ($df = 63$) examined the difference between ratings for each label and those for *contempt*. [b] t test ($df = 63$) examined the difference between ratings for each label and those for *anger*.
* $p < .01$. ** $p < .001$.

spontaneous response, important in its own right, would also help answer some of the questions raised here about the interpretation of available results. For observers in one's own culture, of course, each reader has his or her own response as a guide, but the issue is troubling when observers are from another culture.

Only one method, forced choice, supports the conclusion that across (literate) cultures, a very high proportion of subjects agree on one specific emotion for each of the hypothesized facial expressions. Of course, forced-choice format has its place, but care must be taken in exactly what conclusions are drawn from results gathered with forced choice. For example, the question arises of whether this method has created a false impression that most subjects would spontaneously use the same specific emotion label for a given face.

When subjects were allowed to choose any label they wanted, they did not restrict themselves to synonyms for the 7 ± 2 words used in forced choice, or even to an emotion label. For some types of expression, the recognition scores from free labeling were repeatedly lower than those from forced choice. Forced choice might funnel a range of genuinely different interpretations into one choice, thereby inflating the appearance of consensus. When the same subject responded with both forced-choice and freely chosen labels, the results corresponded in from 89.9% of cases to as low as 7.7%, depending on the expression (Boucher & Carlson, 1980). A similar conclusion came from contrasting the various results across the studies of the "contempt" expression: Recognition scores from forced choice were high; those from free label were low. The free-label format thus could give results different than those given by forced choice, reinforcing the argument earlier that results from forced choice could be specific to that method.

If we are to disregard results from the other methods and to base our conclusions solely on results from forced choice, then we would need a convincing argument that forced choice is indeed the method of choice. Various considerations argue in the opposite direction. One study showed that the forced-choice method could produce apparently conflicting or contradictory results, depending on the experimenter's a priori choice of just which emotions to include on the list of options. In addition, the assumptions underlying the forced-choice format (all-or-none, mutually exclusive options) have not been verified by studies with other formats or by studies of the mental concepts for emotion expressed by the words observers are forced to choose among (Russell, 1989b; Russell & Bullock, 1986). Needless to say, forced choice can be useful in certain circumstances, but conclusions about the judgment of emotion from facial expression cannot rely on this one method alone.

The second most commonly used method has been free label. Most of the data on free label are from Western cultures. In the two non-Western cultures studied, the recognition scores were high for some expressions, moderate for some, and negligible for some. Recognition of the "contempt" and "fear" expressions was probably not significant. Such results might appear to offer at least a partial replication of the forced-choice results. But, like forced choice, free label forces subjects to respond in

terms of emotion and to choose one emotion label as applicable in an either-or manner.

Different results occur with still different response formats. The most open-ended approach has been used infrequently but did suggest that some subjects might not respond in terms of an emotion. Facial expressions might sometimes be interpreted as parts of instrumental action or in terms of the situation. Quantitative ratings help clarify the freely chosen labels, which appear to suggest large individual differences; but quantitative ratings and freely chosen labels together indicate that each observer finds a range of labels applicable to a given face, to different degrees (Russell & Bullock, 1986, pursue this point). The applicable range includes not just synonyms of one basic emotion, but what are claimed to be distinct emotions. In short, as we move from more restrictive to less restrictive attempts to capture the observer's response, a different interpretation of that response emerges. There appears to be little justification for claims that a facial expression is interpreted in terms of one specific emotion category rather than a broad range of overlapping categories.

Lack of Contextual Information

In the studies listed in Tables 2 and 3, observers knew nothing about the expresser except what was shown in the face. They knew nothing about what caused the facial expression, about the setting in which it occurred, or about the expresser's other behavior, past or present. In folk wisdom, this additional information (collectively known as the *expresser's context*) is useful in interpreting the facial expression. Thus, whereas a smile in the context of just having received a gift might be interpreted as a sign of pleasure, a smile in the context of just having spilled soup might be interpreted as a sign of embarrassment, and a smile in the context of greeting an adversary might be interpreted as an act of politeness. To refer to a specific facial expression as a *signal* requires that the expression communicate the hypothesized message not only when the face is seen alone, but when seen embedded in a reasonable range of naturally occurring contexts. This prediction of the universality thesis has received little attention but is important for the hypothesized communication function of facial signals and for the hypothesized evolutionary origin of those signals.

Ekman et al. (1972) reviewed the early studies on this topic but focused on a different question than the one being raised here. They asked the question of dominance: When both the facial expression and the context are known and when the two are discordant (i.e., when face alone and context alone would suggest different emotions), which source of information is the more influential? They found no simple answer to this question: "There seems to be no question that either source, face or context, can, on given occasions, be more salient or more useful or more of a determinant of the combined judgment than the other" (p. 150).

The results of the early studies are more consistent if the question of dominance is set aside and the question of generalizability is asked: Does the emotion attributed to the expresser when his or her face is seen alone generalize to attributions made

when the face is seen in context? The early studies showed consistently that contextual information influenced the observer's judgment (Frijda, 1958; Goldberg, 1951; Goodenough & Tinker, 1931; Munn, 1940; Vinacke, 1949). More recent studies also focused on the question of dominance. But again, if the question of generalizability is asked, then the answer is the same. Although the recent studies found mixed results on the question of dominance, all studies found that contextual information influenced the observers' judgment (Fernandez-Dols, Sierra, & Ruiz-Belda, 1993; Fernandez-Dols, Wallbott, & Sanchez, 1991; Knudsen & Muzekari, 1983; Motley & Camden, 1988; Nakamura, Buck, & Kenny, 1990; Spignesi & Shor, 1981; Wallbott, 1988; S. G. Watson, 1972).

Studies reviewed in the last two paragraphs can provide only a weak indication on the issue of generalizability. Ekman et al. (1972) pointed to various technical ambiguities in the early studies. But for present purposes, a more important problem is that of sampling. The set of contexts studied was not always representative of the contexts that actually surround facial expressions. In some studies, the pairing of face and context was dictated by a factorial design (all faces paired with all contexts) rather than by the likelihood of that pairing occurring in the nonexperimental world. The judgment process invoked by an unusual pairing might not be the same as that invoked by a more usual pairing. The available evidence therefore provides no strong conclusion but does demonstrate that context can influence the judgment of emotion from face and therefore that the research strategy of focusing exclusively on judgments of the face alone is of questionable generality to naturally occurring situations.

Judgment Studies in Isolated Cultures: The Evidence From "Those Associated but Little With Europeans"

> For six emotions—happiness, fear, anger, sadness, disgust, and surprise—universality of facial expression has been established by showing that members of cultures having no visual contact can correctly recognize one another's expressions. (R. Brown, 1986, p. 522)

This quotation from an undergraduate textbook underscores the importance attached to the studies of visually isolated cultures, to which I now turn. All of the studies reviewed here so far suffer from one widely acknowledged problem. As Darwin (1872/1965) noted, the question is what occurs with "those who have associated but little with Europeans" (pp. 14–15). Although non-Western subjects were included in the studies cited in Table 2, all were students who probably had contact with Westerners in person or through films, television, books, drama, magazines, and so on. The degree of this contact is difficult to estimate, but the possibility exists of their having learned some culture-specific Western facial gestures or a Western interpretation of a given facial pattern. This problem of cultural contact motivated the study of people isolated from Western culture, who would provide the most telling answer to this problem.

The evidence on isolated cultures comes from two sources: (a) a series of studies organized and guided by Ekman, Friesen, and Sorenson and carried out from about 1967 to 1970 and (b)

a single study conducted by Boucher and Carlson (1980). The principal methods used, the groups studied, and references are given in Table 14.

The Ekman, Friesen, and Sorenson Studies

This research began when Ekman and Friesen saw films taken by Sorenson and others of two isolated societies in New Guinea (Ekman, 1972, p. 210). The three researchers then traveled to New Guinea and organized a series of studies. The same questions, same concepts, and same stimulus materials reappear in each of the studies, although the studies were conducted with different methods, by different experimenters, and in different places. Thirty-two of the photographs used were published by Sorenson (1975, 1976).

No single account of this entire research program has been published. Lack of a published overview has unfortunate consequences. We do not know the precise original hypothesis: For example, in some studies, contempt and disgust were treated as separate emotions, in others as one emotion. Some of the studies were never published in full but were reported in secondary sources, with neither method nor results fully detailed. Some studies were apparently never reported at all. For instance, Ekman et al. (1969) mentioned that their report concerned data only from the most Westernized Fore subjects and that further data were available from less Westernized subjects—data that were "similar" with the "exception of the sadness category" (p. 88). Apparently these data were never published. These investigators also reported that South Fore facial expressions were shown to South Fore observers, but apparently these data were never published. It is therefore unclear how many studies were conducted altogether, precisely what hypotheses were tested, in some cases exactly what methods were used or results obtained, and how selective the published reports have been.

In this research program, experimental control appears to have been low. Such problems are not surprising in cross-cultural work of this sort, but they are important nonetheless in evaluating the results. Sorenson (1976) commented,

> Neither Ekman nor Friesen, who participated in this part of the study, knew Melanesian-Pidgen or Fore. My own Melanesian-Pidgen was good; but I was not up to following native discourse or making myself understood in Fore beyond the simplest messages. Therefore, for the least acculturated Fore we had to rely on Fore translator–assistants to explain the task and to relate the stimulus stories. We were not able to monitor this communication. In view of the Fore communicational conventions, it was likely that at least some responses were influenced by feedback between translator and subject. The Fore, even those trying to be most Western, could not be expected to have internalized our Western concepts of a testing situation sufficiently to avoid "leaking" information. The suggestion that free exchange of information was "cheating" was quite incomprehensible to the Fore and alien to their view of language as an element of cooperative interaction among close associates. The best we could do was to impress our assistants continually with the importance of not discussing the pictures with the subjects, and particularly, not telling them which pictures to select or suggesting key features for which to watch. The effect of our cautions could not be determined. . . .
> Both native assistants and subjects were generally very eager to do things in the Western way. Like other things Western, our pic-

Table 14
Five Methods Used With Isolated Cultures

Method	Culture	Reference
Standard	Sadong of Borneo	Ekman, Sorenson, and Friesen (1969)
	Fore of New Guinea	Ekman, Sorenson, and Friesen (1969)
Free labels	Fore of New Guinea	Sorenson (1975, 1976)
	Bahinemo	Sorenson (1975, 1976)
Dashiell	Fore of New Guinea	Ekman and Friesen (1971)
		Sorenson (1975, 1976)
	Grand Valley Dani	Unpublished, reported by Ekman (1972)[a]
Posing emotions	Fore of New Guinea	Unpublished, reported by Ekman (1972)
Selecting photographs	Temuan of Malaysia	Boucher and Carlson (1980)

[a] In a footnote, Ekman et al. (1972) reported, "In a study of another preliterate culture in New Guinea (the Dani), which is considerably more isolated than the Fore, essentially the same results were found (Ekman, 1972; Ekman, Heider, Friesen, & Heider, in preparation)" (p. 162).

tures and procedures were the subject of considerable interest and active discussion by the behaviorally alert Fore. They were quick to seize on the subtlest cues for an indication of how they should respond and react. This undoubtedly skewed our results. (Sorenson, 1976, pp. 139–140)

Four principal methods were used to gather data from four separate cultures, as outlined in Table 14. I consider each method in turn.

Standard Method

Two of the reported studies (Ekman et al., 1969) used the standard method already discussed: Subjects were shown preselected still photographs largely of posed facial expressions and then were asked to choose one of six alternatives. The results are summarized in the left half of Table 15. The three cases involving facial expressions of "happiness" obtained high agreement with prediction (82% to 99%)—the subjects apparently understood the instructions and could perform the task. Of the remaining 15 cases, modal responses were as predicted in 12, although the proportion of subjects agreeing with prediction was greater than 50% in only 6. The range was from 29% to 56% for Pidgin-speaking Fore, from less than 19% to 54% for the Fore-speaking Fore, and from less than 23% to 64% for the Sadong. Statistical tests compared obtained results with completely random selection of labels, which would result in 16.7% agreement. No statistical test was offered to show that results for negative emotions were reliably greater than what would be

Table 15
Percentage of Agreement With Prediction for Isolated Cultures

	Standard method			Dashiell method	
	Fore culture		Sadong culture	Fore culture[a]	Dani culture
Facial expression	Pidgin	Fore	Bidayuh	Fore	Dani
n	18	14	15	130	34
"Happiness"	99	82	92	92	98
"Fear"	46	54	40	88[b]	80[b]
"Anger"	56	50	64	90	68[b]
"Disgust/contempt" [c]	29	44	<23[d]	85	91
"Surprise" [e]	38	<19[d]	36	98	89
"Sadness"	55	—[d]	52	81	77

Note. For the standard method, there were six response options; 16.7% was the level that was due to chance.
[a] This was a group of Fore children. Similar results occurred with Fore adults. [b] Median of values given by Ekman (1972). [c] There is no Neo-Melanesian Pidgin word for disgust or contempt; therefore, a phrase was used: "looking at something which stinks." [d] Modal choice was another emotion, and no figure of percentage correct was given, although it was less than the figure given. [e] There is no Neo-Melanesian Pidgin word for surprise; therefore, a phrase was used: "looking at something new."

Table 16
*Freely Chosen Modal Label and Percentage to Facial Expressions of Emotion,
From Sorenson (1976)*

Facial expression	Fore-speaking Fore					
	Most contact		Intermediate contact		Least contact	
	Label	%	Label	%	Label	%
"Happiness"	*Happiness*	82	*Happiness*	67	*Happiness*	60
"Surprise"	Fear	45	Fear	31	Happiness[a]	31
"Sadness"	Anger	56	Anger	57	Anger	53
"Anger	*Anger*	50	*Anger*	49	*Anger*	48
"Fear"	*Fear*	54	*Fear*	30	*Fear*	23
"Contempt"	Anger[b]	36	*Contempt*	28	Anger	27
"Disgust"	Contempt	39	Happiness	24	Happiness	23

Note. Response, when as predicted, is underlined.
[a] The second most common response was the one predicted, *surprise* (26%).   [b] The second most common response was the one predicted, *contempt* (29%).

expected by random selection among negative labels (25%). The Fore-speaking Fore chose *anger* for the "sad" faces 56% of the time, a greater degree of agreement than for any other set of faces except "happiness."

Overall, with the notable exception of happiness, the results were less in agreement with prediction than were those obtained from nonisolated, literate cultures, summarized earlier in Table 2. Of course, the illiterate and isolated subjects were unfamiliar with being questioned and with the idea of a scientific experiment. Even if the 12 of 15 cases for nonpositive emotions did yield results greater than chance, the results must be viewed in light of the problems already described with the standard method and the likelihood "that at least some responses were influenced by feedback between translator and subject" (Sorenson, 1976, pp. 139-140).

Another fact to remember about these two studies is that the subjects "had extensive contact with Western culture" (Ekman & Friesen, 1971, p. 125). The Fore for whom data were reported were the "most Westernized" (Ekman et al., 1969, p. 87). They were "most influenced by contacts with Westerners (government, missionaries, and others)" (p. 87). Of the Sadong, "many had seen a few movies" (p. 88). As Ekman and Friesen (1971) later acknowledged:

> because all the cultures [that Ekman et al., 1969] compared had exposure to some of the same mass media portrayals of facial behavior, members of these cultures might have learned to recognize the same set of conventions, or become familiar with each other's different facial behavior. (p. 125)

Freely Chosen Labels

Freely chosen labels were reported for two groups by Sorenson (1975, 1976): 100 Fore and 71 Bahinemo (who were more isolated than the Fore).[11] The Bahinemo showed no consistent association between facial expression and emotion label, except for a tendency to see all the (Caucasian) faces as angry.

Separate results were reported for those Fore with most, intermediate, and least contact with Western culture; and a summary is provided in Table 16. The data from the Fore do not always allow calculation of the percentage agreeing with prediction but do give some indication. For "happiness," "anger," and "fear" expressions, the modal response was as predicted; consequently for these three, the median recognition score is available: 67% for "happiness," 49% for "anger," and 30% for "fear" expression. Figures are also available for two of the three groups for the "contempt" expression; therefore 28% would be the median. For the "surprise" expression, one group selected *surprise* 26% of the time; the median would be somewhat less. For "sadness" and "disgust," neither the first nor second choice was as predicted; percentage correct would therefore be less than the figures shown in Table 16. In addition, a clear violation of prediction occurred when a majority in each of the three Fore groups labeled the "sad" expressions as *anger.* (Recall that the majority of the Fore-speaking Fore chose *anger* for the "sad" expression in the forced-choice method as well.)

Overall, agreement with prediction among Fore was less than in free-label data in less isolated cultures. It is also not known how subjects' actual responses were clustered together to produce the results shown in Table 16. In the light of previous reports of freely chosen labels, I doubted that the Fore observers had limited themselves to precisely six labels. When asked this question, Sorenson (personal communication, July 6, 1991) replied that the subjects produced "hundreds of words . . . stories and all kinds of things." Presumably, the results of Table 16 indicate the amount of agreement obtained on fairly broad clusters of responses.

---

[11] Sorenson (1975, 1976) also presented results from 15 members of the Sadong and seemed to imply that the response measure was a freely chosen label. However, the same results had been presented earlier, and it was clear that the response measure was forced choice (Ekman, Sorenson, & Friesen, 1969).

The Dashiell Method

Because of disappointing results produced so far, the researchers turned to a method introduced by Dashiell, in which each emotion was represented by a story. On the trial for happiness, for instance, the subject was told that someone had just met a friend and was happy. The subject was then asked to select a facial expression for the story protagonist. A summary of the results is shown in the right half of Table 15. Much higher recognition scores were obtained with the Dashiell method than had been obtained earlier with the standard method or with free labeling. It is especially unfortunate that these studies have not been published in full. The Dashiell method has most of the features of the standard method, but has some unique features as well, which are worth considering in detail.

First, these experiments were conducted in the field rather than in a laboratory, and it is unlikely that the experimenters were able to control precisely the information given to the subject. For example, it is difficult to tell an emotional story without altering one's tone of voice and facial expression. The kind of leakage described by Sorenson (1975) might have similarly played a role.

Second, the response measure reduced the number of choices: All Dani subjects and all Fore children were given two, and all Fore adults three, photographs from which to choose. Random response selection would therefore produce 50% or 33% correct, respectively. The results reported with this measure were not corrected for guessing. An even higher level could sometimes be produced by simply distinguishing positive from negative responses. If a sad story was given to a Fore adult, for example, and the choice was among three options, a "sad," an "angry," and a "happy" facial expression, then 50% recognition would be achieved by distinguishing positive from negative expressions (eliminating the "happy" expression and choosing randomly between the other two).

The overall results also depended on just which faces subjects were shown on a given trial, and not all possible pairs or triads were included. Indeed, some important ones were omitted. For the Fore sample given the disgust story, an "anger" expression was never one of the possible choices. For the Fore children given the fear story, a "surprise" expression was never a possible choice; for those given the anger story, a "disgust" expression was never one of the possible choices. Therefore, the precise alternatives the subjects were given must be considered. For example, the median recognition score given in Table 15 for the Dani responses to the anger story (68%) derived from four judgments: anger–happiness 94%, anger–fear 76%, anger–sadness 61%, and anger–disgust 48%, where random choice would be expected to yield 50%. If we set aside the anger–happiness result because it could be based on a simple positive–negative discrimination, then the median would be 61% (22% when corrected for guessing). This apparent ability to associate the anger story with the predicted face varied (whether reliably is not known) with the alternative. Similar examination of specific alternatives for the fear story showed that Fore adults were as or more likely to choose the "surprise" expression as the "fear" expression. (Fore children were not asked to make this discrimination.)

Third, over the course of the experiment, subjects could have learned to associate faces and stories in something like a learning-to-learn procedure. Each subject participated in a number of trials, and for each trial for a given emotion story, exactly one of the expression options recurred, namely, the "correct" expression. Some subjects might have noticed which kind of facial expression always appeared with a given story. For example, suppose that on the first trial a Dani hears the story about disgust (the protagonist smells something bad) and is shown two pictures, a frown and a wrinkled nose. Suppose further that this subject has no idea which facial expression to choose for the disgust story, and so chooses randomly. Now, on the next trial that concerns the disgust story, the wrinkled nose reappears, this time paired with a different alternative. Whatever alternative is offered, that subject, eager to accommodate the experimenter, might remember the wrinkled nose from the first trial. If so, that subject will have learned that the wrinkled nose goes with the disgust story. Moreover, now suppose that the next trial concerns, say, the surprise story and that the choice is between the "surprise" expression and the wrinkled nose of "disgust." Our subject might now guess that the wrinkled nose is not correct for this story. Thus, he or she could select the other (appropriate) face. On later trials for other emotions, this same subject could begin to guess the correct facial expression by eliminating alternatives previously associated with other stories. Of course, no evidence shows that anything like this fictional reconstruction actually occurred. But the possibility that some such learning-to-learn could have occurred on some trials for some types of facial expression for some subjects cannot be ruled out. The appropriate control conditions were not included that would allow us to do so.

Fourth, the story and the emotion were confounded. For happiness, it is unclear whether the Fore or the Dani subject selected the smile as indicating happiness or as the response to meeting a friend or whether they selected the wrinkled nose as indicating disgust or as just a response to foul odors. Thus, the story alone, with no emotion presented or inferred by the subject, might have accounted for or contributed to the selection of a facial expression. Ekman's (1980) photographs of the Fore show various facial movements, but the question is whether the Fore interpreted these movements as expressions of emotion.

Imagine you are the subject in this experiment and that you know nothing about emotions. You are told the "anger" story concerning a protagonist about to fight. You might reasonably select the face with glaring eyes and set jaw. You are told a "surprise" story concerning someone looking at something new and unexpected. You might reasonably select the face staring at something. You are told a "fear" story concerning someone who sees a dangerous wild pig. You might reasonably pick a face that is tense, or where the eyes are staring, or where the mouth appears about to yell. With the data available, it is impossible to disentangle face-emotion knowledge from face-situation knowledge. Conceivably, the results might have little or nothing to do with emotion per se.

It is undoubtedly informative that people in isolated, illiterate cultures associate the same facial movements with greeting a friend, fighting, and smelling something bad, as do we in the

West. The question is how to interpret the information. It might be argued that the only plausible interpretation is in terms of emotion. If these stories captured pancultural antecedents of emotion, as specified a priori by the experimenter's theory (and no other methodological problems occurred), or if we knew that the Fore associated each story with a specific emotion, then emotion would remain a compelling interpretation. Unfortunately, this line of defense is undermined by the way in which the stories were developed in prestudies (Fernandez-Dols, in press): "Simple stories were developed within each culture to be relevant to that culture" (Ekman, 1972, p. 271), and emotion played no role in generating the stories. For example, in the prestudy of the Fore, some members were shown the photographs of facial expressions and were asked to make up a story about the person shown. The subjects were shown a smile, and they said that the smiling person was greeting a friend. They were shown a wrinkled nose, and they said that the person was smelling something bad. Later, in the actual study, this process was simply reversed. That is, other Fore were later told the story of someone greeting a friend and were asked to select the face most associated with that story; they selected the smile. When told the story of someone smelling something bad, they selected the wrinkled nose. In neither phase was emotion necessarily involved.

We Westerners seem to infer readily that happiness must have been involved in the association between the smile and greeting a friend or that disgust was involved in the association between smelling and wrinkling the nose. But we have no evidence that it was something that the Fore inferred or even thought about. We Westerners find it plausible to suppose that the Fore think in terms of happiness, disgust, and other emotions, but the purpose of this study was to gather evidence on this question. Peoples of other cultures speaking other languages may conceptualize psychological states somewhat differently than is done in Indo-European languages (for a review, see Russell, 1991b). Some cultures do not have a word equivalent to *emotion,* nor do they appear to categorize emotions in the same way as is done in English.

The Dashiell method creates several uncertainties. Some of these uncertainties arise from the features it shares with the standard method. Some arise from its unique features and field setting. Each of its associated problems separately could have inflated the amount of agreement with prediction. When the various problems all co-occur in the same study, it is impossible to estimate their joint influence. The within-subject design of this experiment coupled with the small number of emotions studied (and perhaps the kind of leaking described by Sorenson, 1976) may have provided subjects with clues to solving a puzzle. Even a partial solution could have inflated the percentage of agreement. It is therefore impossible to know how much, if at all, the obtained results were due to subjects' associating emotions with facial expressions.

Posing Expressions

In a review of the literature, Ekman (1973) referred to a Study 2 in Ekman and Friesen (1971). Ekman and Friesen

(1971) reported this study in a single sentence: "Visually isolated members of the South Fore posed emotions, and college students in the United States accurately judged the emotion intended from their videotaped facial behavior" (p. 128). Elsewhere, Ekman (1972) gave more detail. The videotapes were of 9 members of the Fore (the number taped was not specified). The American judges were 34 university students, each of whom viewed videotapes of all 9 Fore and apparently chose one emotion label from a list of six. The recognition scores were as follows: happiness 73%, sadness 68%, anger 51%, disgust 46%, surprise 27%, and fear 18%. The last two figures did not differ significantly from those expected by chance.

As in the Dashiell Method, the Fore were given stories to help them pose the facial expressions: A wrinkled nose was produced in response to the story of smelling something bad. Once again, it was unclear whether the Fore could pose a facial expression for disgust per se or could simply show what happened when they smelled something bad. The judgment task given the American subjects was within subject and used a forced-choice response format, with all the problems entailed.

The Boucher and Carlson Study: Selecting Photographs

Boucher and Carlson (1980) studied 31 Temuans, a forest-dwelling group of proto-Malay aborigines. "While these people are by no means the cultural isolates studied by Ekman in New Guinea, their contact with Western culture is quite limited" (p. 276). Each Temuan was shown 25 photographs of five American men, each posing from four to six facial displays that the authors believed expressive of basic emotions. The subject first saw all photographs (four to six) posed by a single actor. The subject was asked to choose the one photo where the American poser looked, say, angry. The subject was then shown the next poser and asked to select a different emotion, and so on through the five posers. This procedure was repeated until the subject had selected a photograph for each emotion available from each poser. The observer was not asked to select any emotion not present for a given poser.

On trials in which the subject was asked to select an "expression of happiness," 90.3% of responses were as predicted. The corresponding figures were 84.5% for "disgust," 79.8% for "surprise," 74.0% for "sadness," 72.9% for "anger," and 53.2% for "fear." Each of these figures exceeded chance to a reliable degree.

The facial stimuli were posed and preselected, the design was within subject, and each subject was shown the entire set of stimuli at least five times. Although the usual role of word (e.g., *happy*) and photograph (e.g., smile) was reversed in this study, the method was again forced choice, this time between four to six alternative faces. The method would therefore be subject to some of the same problems as created by the standard method, including forced-choice response format. In addition, the number of trials for each emotion varied from two (because only two posers posed "fear") to six (because all six posers posed "anger," "disgust," and "happiness"). Thus, the subject was asked to select fear on only two occasions, namely, those in which the hypothesized fear face was present. Some subjects might have no-

ticed this fact. If so, then this clue might have allowed those subjects to eliminate the "fear" alternative when another emotion was the target. Suppose, then, that on one trial the subject was asked to select the angry person among four alternatives but could eliminate the "fear" and "happiness" faces. Then the number of alternatives actually relevant would be half the nominal number. It was, therefore, difficult to estimate actual base-rate probabilities.

Conclusion

Recognizing that studies of literate societies in the modern world of mass communication could not answer the question of universality, Ekman et al. (1969) turned to isolated, illiterate societies. Ekman et al. (1969) hypothesized that sadness, like other basic emotions, had a biologically based, universally recognized facial signal. Photographs of the hypothesized facial signal for sadness were shown to Fore-speaking observers reared in a culture relatively isolated from outside influence. When asked what emotion the person with the "sad" expression was feeling, over 50% replied *anger*. When shown the "sad" expression and asked more directly whether the person was happy, surprised, afraid, angry, disgusted, or sad, 56% said *angry*. When told a story about someone who was angry and about to fight and shown two photographs, the hypothesized anger expression and the hypothesized sad expression, 61% selected the anger expression, and 39% selected the sad expression. This pattern of results from the Fore is not easy to reconcile with the original hypothesis.

Even if the pattern of results were more clear-cut, the methods used with isolated societies left many uncertainties. Although several different methods were used, some design features remained the same. For example, most studies used the same still photographs of mainly posed facial expressions. In most cases, observers provided their response in some sort of forced-choice format, either choosing among words or photographs. All were within-subject designs. No information was provided on whether the facial stimuli that were used corresponded to naturally occurring facial expressions in the culture studied. No information was provided that the emotion response categories corresponded to indigenous concepts. The lack of an indigenous word for surprise among the Fore reinforced this concern.

The easiest comparison across cultures is with studies using the same method. In the isolated cultures, the standard method showed a nonrandom association between some facial expressions and emotion labels. Nevertheless, because the particular subjects had at least some contact with Westerners and Western mass media, because of the problems associated with the standard method, and because of poor experimental control, these studies could not provide a strong test of the experimenter's hypothesis. These technical problems aside, the specific hypothesis that happiness, surprise, fear, anger, disgust, and sadness were highly recognizable from facial expression fit the data clearly only for happiness. This negative result has not received the attention it deserves. The data on happiness demonstrated that the subjects understood the task per se, which in any case

was quite simple. Highly selected intense facial expressions were shown one at a time. A list of six emotion names was read to each subject. Each name on the list supposedly represented one discrete, basic, universal emotion. The list remained the same for every photo and was "repeatedly read to them after each photo" (Ekman, 1973, p. 210). Other than for *happiness*, the emotion label was not associated with specific facial expressions in the manner and to the degree predicted.

Studies of freely chosen labels did not produce much support either. One group, the Bahinemo, showed no consistent pattern except a tendency to label all facial expressions as *anger*. The other group, the Fore, gave partial agreement with prediction. How their free labels were clustered together was not specified, and therefore no strong conclusions are possible.

The researchers then changed methods. The results that agreed most with prediction (Boucher & Carlson, 1980) were apparently obtained with the most Westernized subjects, and their method had problems associated with within-subject design, preselected, posed, still photographs, and forced-choiced response format, as well as potential internal problems of its own. The other results that agreed with prediction were gathered with the Dashiell method. This method had the most serious problems, the combination of leakage from translator to subject, a design that allowed learning to learn, and a confounding of emotion with story. Moreover, the story had been generated within the culture in response to the facial expression. Even so, fear was not distinguished from surprise, nor anger from disgust. Technical problems aside, this last result indicated that if the Fore and Dani interpreted facial expressions in terms of emotions, they might have done so in terms of broader clusters than would be indicated by the English words *surprise, fear, anger,* and *disgust*.

The focus of these studies was on agreement with an a priori hypothesis, but some unexpected but important results were also obtained, such as the Fore labeling the "sad" photographs as *anger*, both in the free-label and in the standard forced-choice task. Because this result was not anticipated, it was less likely to be accounted for by experimental demand or help from the translator. None of the studies were aimed at revealing the role of culture in interpreting facial expressions—quite the contrary. Nevertheless, such evidence suggests that culture does play some role in the interpretation of facial expressions.

The data on isolated cultures are also consistent with the hypothesis that degree of contact with Western culture is directly related to the amount of agreement with the predictions of the Western investigators. The most straightforward comparison relies on studies that were conducted with the standard method. A clear trend is shown in Table 17 for all types of facial expression except "happy." Although the college students who served as subjects might have obscured cultural differences, the amount of agreement with prediction from non-Western literate cultures was reliably lower than that from Western cultures. In turn, more isolated cultures provided even weaker agreement with prediction. And this decrement occurred despite the softening of methodological control mentioned by Sorenson (1976) to accommodate illiterate subjects unfamiliar with experimental procedures. In addition, only among the iso-

Table 17
Median Percentage Agreement for Forced Choice

Culture group	Facial expression					
	"Happy"	"Surprise"	"Sadness"	"Fear"	"Disgust"	"Anger"
Western	96.4	87.5	80.5	77.5	82.6	81.2
Non-Western literate	89.2	79.2	76.0	65.0	65.0	63.0
Illiterate, isolated	92.0	36.0	52.0	46.0	29.0	56.0

Note. Figures are median values from Tables 2 and 15 for results gathered with the standard method. The number of groups was 20 for Western, 11 for non-Western, and 3 for illiterate, isolated cultures.

lated cultures were there strong contradictions of the experimenter's predictions, as when the Fore consistently labeled the "sad" expression as *anger.*

Further evidence for the same trend can be seen across the studies of isolated cultures. The greatest agreement with prediction was obtained by Boucher and Carlson (1980), whose subjects were apparently the most Westernized. Among the groups given the standard method, Pidgin-speaking Fore were the only group to produce above-chance agreement on all six types of facial expression, and Pidgin is a language created for communication between cultures. In the studies of freely chosen labels, again greater agreement with prediction came from the Fore, who had more contact with Westerners, and less agreement—indeed no agreement—with prediction came from the Bahimeno, who had less contact with Westerners.

With the important exception of the Ducci et al. (1982) study, the comparisons just offered to suggest the role of culture were after-the-fact interpretations. Cultural contact was confounded with education, language, and familiarity with experimental procedures. On the other hand, the cultural differences in Tables 2 and 3 were reliable, and the emergence of cultural variation in a set of studies specifically designed to demonstrate universality should encourage researchers to follow Ducci et al.'s important lead by exploring the role of culture directly. If cultural variations are reliable, then the question arises of how people in non-Western cultures interpret facial movements.

Validity of the Judgment Studies

I now turn to a summary evaluation of the evidence. Do the studies show what previous reviews have claimed they show? I consider ecological, convergent, and internal validity, although the concerns are somewhat overlapping.

Ecological Validity

Little information is available on what facial expressions occur naturally in the various societies studied and in what frequency and in what contexts. Little information is available on what interpretations observers naturally make of the facial expressions they actually encounter and in what contexts those interpretations occur. It seems unlikely that the facial expressions used as stimuli, the limited options sometimes given subjects for making their judgments, and the experimental contexts

in which judgments were made were representative samples of naturally occurring stimuli, responses, and judgmental contexts.

Even if the results reviewed in this article were internally valid, they would not necessarily tell us what occurs in everyday situations. Preselected, posed facial expressions, forced-choice response format, within-subject design, and lack of contextual information are potential challenges to ecological validity. For example, with a forced-choice response format under experimental conditions, subjects might associate soft hair with cowardice or coarse hair with bravery, as suggested by Aristotle. Even if they did, that result would not tell us that when soft hair is encountered in everyday events, it is noticed or so coded. Soft hair would be embedded in many other features, especially behavior, which might so overwhelm the judgment that the effect of soft hair would be negligible. To return to faces, suppose that facial expressions rarely or seldom occur in the intensity and clarity portrayed in the photographs shown to subjects or that observers rarely or seldom have the time to study the face as they did when shown a still photograph. Then, their degree of recognition of specific emotions from the kind of photographs used in the experimental context might not tell us what happens in most everyday face-to-face encounters.[12] This concern was

---

[12] Researchers had their reasons, of course, for not attending to ecological questions. For example, facial stimuli were posed and carefully selected because natural expressions were said to be often inhibited or modulated by display rules (Ekman, 1972; Izard & Malatesta, 1987). The word *often* raises questions. If "pure" (uninhibited and unmodulated) expressions are rare but nevertheless do occur, then it still should be possible to obtain photographs of them for scientific purposes. So, the question is, why not? In the extreme case that uninhibited, unmodulated expressions are so rare that no photographs are possible, then the question is, Do people ever see the pure facial expressions? If not, how on Ekman and Friesen's (1975) theory, do children acquire the ability to associate emotions and faces? And what is the purpose of a theory about a phenomenon that so rarely occurs? Izard (1977) argued that at least infants show the pure facial expressions, but this argument was disputed by Oster et al. (1992) and Camras (1992). It was also argued that still photographs fail to capture the full impact of a facial expression because actual facial expressions are moving. Thus, exaggerated poses merely compensate for having still rather than moving expressions. One problem with this solution is that a moving expression might be interpreted not just more easily, but in a different manner than still expressions. In any case, the solution is not to compensate for one prob-

reinforced by the decline in recognition when features of method were altered.

The preceding comments carried no implication of criticism of the original researchers. Providing such information was not their purpose. However, I want to argue that ecological validity must be or must become a concern. I have three reasons in mind. First, as is evident in Table 1, reviewers of this literature have not limited their conclusions to the experimental context (see also footnote 2). Apparently, reviewers assumed the available evidence to be ecologically valid. Writers in this field have not sufficiently emphasized the lack of available information concerning ecological questions. For example, in summarizing their findings for a general audience, Ekman and Friesen (1975) wrote,

> This book is about faces and feelings—your own and those of the people around you. The first focus is on what the feelings look like, in other people's faces and in your own. Photographs show the facial blueprints of the major emotions. . . . You can use this information about the blueprints of facial expression to better understand the feelings of others, even when they are trying not to reveal their feelings. Or you can use the knowledge of the blueprints of facial expression to learn about your own face, to become more aware of what your face is telling you about how you feel and what your face is telling others. . . . This book is for psychotherapists, ministers, physicians and nurses, trial lawyers, personnel managers, salesmen, teachers, actors. . . . The psychotherapist . . . must be alert to what the face may tell him about his patient's feelings. The patient's face may show the emotion being experienced even when it is too upsetting to put into words. . . . The trial lawyer often can't trust the words of a witness or client. He needs another source, such as the face, to tell him how the person really feels. (pp. 1–3)

Ekman (1989) later pointed out that "there are no data about how many expressions for each emotion are universal. Nor is it known how often these universal expressions of emotion are seen in ordinary social life" (p. 154).

The second reason concerns those theories of emotion that have been said to be supported by this evidence. The theories of Ekman (1972), Izard (1971), and Tomkins (1962–1963) consider the advantages of an emotion-signaling system so great as to have affected the course of human evolution. This aspect of their theories implies that universal facial expressions occur and are recognized in natural settings (or at least were in the past).

Ekman and Friesen (1975) wrote that in every generation children must learn anew the emotional meaning of each facial expression:

> No one taught you how to read those signals . . . you were not born with the knowledge. You have to pick it up. . . . [When you follow] the rules for translating a particular set of facial wrinkles into the judgment that a person is angry, afraid, etc., you do so automatically, on the basis of habits established long ago. (pp. 7–8)

This mechanism of acquiring the ability to recognize emotions

from facial expressions presupposes that each child is exposed to enough examples of the facial expressions and sufficient cues to the associated emotion.

My final reason for emphasizing ecological validity is that interpretation of the kind of results reviewed here may depend on implicit assumptions about ecological distribution. Let me illustrate this dependency with an imaginary study on perception of height. The stimulus set consists of 15 men: 3 men 3 ft, 6 in. tall; 3 men 5 ft., 1 in. tall; 3 men 5 ft, 10 in. tall; 3 men 6 ft, 5 in tall; and 3 men 7 ft., 2 in. Every subject responds to the full stimulus set. Responses are gathered with a forced-choice format with the options *midget, short, average, tall,* and *giant.* I would guess that subjects would assign the 15 stimuli to the five response options with very high agreement and that cross-cultural studies would show high if not perfect agreement. If so, the results would then appear to show five discrete specific categories of height, easily and universally recognized. The results would appear to refute the dimensional view of a single continuum from *short* to *tall.* Of course, no one would design or carry out such a study. If carried out, the study would not be published. If published, it would not be believed. Readers would notice that the method had produced a misleading result. The ecological distribution of height would be violated; the response options would be designed to fit the stimulus categories. No one would find the study valid because everyone would know that height was distributed as a continuum, not as five discrete categories. In this study of height, the flaws of method are obvious, because we know the ecological distribution. Of course, emotions and facial configurations do not form one simple, obvious continuum, and this imaginary study is too simple when applied to the ecology of emotion. The implicit assumptions in the design of the studies reviewed here might be correct—or they might not be, especially as we move to more distant cultures.

To summarize, the studies reviewed in this article were not designed with the question of ecological validity in mind, and those who advocated the universality thesis did not always emphasize its ecological assumptions and implications. Nevertheless, the universality thesis is ultimately about what happens in natural situations. Moreover, inferences drawn from the available evidence and interpretation of that evidence rest on assumptions about the ecological distribution of expressions and the spontaneous interpretation of those expressions. Therefore, ecological questions must be raised. On most such questions, little or no information is available. What little information we have on spontaneous expressions, on unselected expressions, on freely given responses, and on judgments of the face in context raise the possibility that results from the standard method may not tell us what happens in natural settings. This possibility exists in all cultures, but is especially troubling in less familiar cultures.

Convergent Validity

The studies of Table 2 as well as of Ekman and Friesen (1986) and Ekman and Heider (1988) used methods similar enough to define a standard method. The similarity in method in the stud-

---

ies cited raises the question of generalizability to other methods. From responses of college students, we cannot necessarily infer how other samples would respond. From judges' responses to the forced-choice list, we cannot necessarily infer their responses gathered in other ways.

What's more important, there are good reasons to question the generalizability of results gathered with the standard method. Student subjects in more recent studies might have been familiar with the universality thesis. Previewing and the within-subject design allow subjects to compare and contrast the stimuli. The forced-choice response format has been shown capable of producing the appearance of consensus even in the absence of the predicted emotion category.

Some degree of convergent validity has been demonstrated. The various studies listed in Table 2 did use slightly different procedures, stimuli, and lists of response options. For some facial expressions, freely generated labels produced results near those produced by forced choice and therefore demonstrated convergent validity across these two response formats. However, this result was limited to some types of facial expression and largely to Western cultures; for other expressions, convergence was reduced or even failed. Two studies used the Dashiell method (in which a story is substituted for the emotion label; Ekman & Friesen, 1971). And Boucher and Carlson (1980) reversed the standard procedure and asked observers to select a photograph in response to the emotion label. Although the Dashiell method and Boucher and Carlson's method introduce their own problems of interpretation, their use is also a step toward demonstrating convergent validity.

In the studies discussed above, one element of the standard method was altered, but other elements generally remained. For example, in the studies of freely chosen labels, the design was within subject, the stimuli were preselected posed expressions, and so on. Several studies are available on what happens when two elements of the standard method are altered. First, when spontaneous rather than posed expressions were studied and when all of the expressions rather than preselected ones were studied, no convincing evidence for anything beyond a positive–negative discrimination was found. Second, Knudsen and Muzekari (1983) examined the combination of context and forced choice. They allowed subjects an option: Subjects could choose from the usual list of emotions, or they could write in their own response. Half of the subjects were given the usual no-context standard procedure. The other half were given a one-sentence description of the expresser's context. These latter subjects were significantly more likely to choose the write-in option. Thus forced-choice may be less adequate when context is known. Third, Motley and Camden (1988) used spontaneous rather than posed facial expressions in a study on the relative influence of facial and contextual information. Unlike other studies on this question, here contextual information dominated in every case. In other words, lack of contextual information combined with posed facial expression may produce a unique result.

To summarize, convergent validity requires that any proposition be supported by results from an array of methods. Some convergence has been demonstrated, but the question of con-

vergent validity is not really what results obtain when the standard method is kept intact except for one or two elements. Rather, the question is whether the same results occur across a broad range of methods. Little evidence of this sort is available. Generally less supportive results occur in studies in which even one feature of the standard method is altered. Moreover, the results are consistent with lower recognition and recognition in terms of broader categories. Thus, the conclusion of easy recognition in terms of specific emotion categories currently rests on what I call the *standard method*. The validity of that method is the next topic.

Internal Validity

It might be argued that the evidence reviewed here was like much of the evidence available in psychology today. Questions of ecological and convergent validity can be raised, but they should not be allowed to undermine the importance of what has been learned. It might be argued that the available studies on the universality thesis demonstrated competence rather than performance: what observers could do, even if not what they actually do in natural settings or even in other experimental settings. If valid, such a demonstration (especially across a range of cultures) would be of theoretical importance, because it would reveal a human capacity. That such a demonstration was the goal of Izard's (1971) studies could be infered from his explanation for using preselected, posed facial stimuli. Indeed, some points raised here so far (such as preselection of stimuli, posed expressions, and lack of contextual information) did not directly bear on the internal validity of such a demonstration.

Potential threats to internal validity varied from study to study. They included the instructions to subjects, the forced-choice response format, previewing, the within-subject design, the possibility that subjects were familiar with or trained in the experimental hypothesis, confounds, feedback between translator and subject, learning to learn, and experimenter expectancy.[13] Still, I would not argue that any one of these features alone provided a complete alternative explanation of the full array of results seen here.

A more general problem arose from the combination of several of the standard design elements that recurred in many of the studies. For example, the forced-choice response format combined with the stimulus set and the within-subject design might alert the subject to what was expected. Preselection of posed facial expressions created a special stimulus set: maximally similar within types of expression and maximally discriminable between types. The within-subject design then invited subjects to notice the composition of that full stimulus set. In the standard method, the subject could easily notice if, for

---

[13] The articles provided no discussion of any precautions taken regarding experimenter expectancy or demand effects and no debriefing of subjects on the possibility of such effects. Ekman, O'Sullivan, and Matsumoto (1991b) recently commented that they "always thoroughly explained the task and then encouraged subjects to ask questions about the task before proceeding. . . . We have found that subjects better understand what is expected of them after trying it a few times" (p. 294).

example, there were three nearly identical wrinkled noses, three beaming smiles, three glaring frowns, three sneers, and so on— or whatever number of each type was used. Whatever the number of types of expression, the forced-choice response format then typically provided a corresponding number of emotion labels; if seven types of facial expression were shown, then generally seven alternative emotion labels were provided. And judges attempt to use each response category option equally often (Parducci, 1965).

A greater threat to internal validity might have come from a simpler mechanism. No one feature of method or combination of features need be a fatal flaw; rather, various features all pushing in the same direction could have had a cumulative impact. Suppose that each element of the standard method added only a small increment to the total recognition score, each element nudging the observer in the direction of the experimental hypothesis. That is, suppose that having a within-subject rather than between-subjects design increased the recognition score by a few percentage points. Assume the same was true for previewing, fixed order of presentation, experimenter expectancy, instructions to observers, and forced-choice response format. As already described, what evidence we have suggested that some of these method factors might have had a large impact on recognition scores, but suppose that the impact of each element was only modest. If so, as we subtract the influence of each method factor in turn, the recognition score would decline. Until more evidence is gathered, one can only guess what the final result would be.

For such a cumulative effect to be worrisome, it need not have influenced results for every facial expression to the same degree. For example, suppose that everyone effortlessly and spontaneously labeled the "happy" expression as happy. If so, changes in method would leave its recognition score high. Similarly, suppose that everyone recognized an unpleasant expression. But would they discriminate within the set of negative expressions—"anger," "disgust," "fear," and "sadness"? For these four expressions, the median recognition score from the standard method was about 82% for Western literate cultures, about 65% for non-Western literate cultures, and about 49% for the isolated, illiterate cultures. All of these figures were well above the level (25%) that could be achieved by choosing randomly among these four labels. Nevertheless, replacing the forced-choice response format with free label was generally found to lower recognition scores. Replacing the within-subject design with a between-subjects design lowered recognition scores. What would be the recognition scores for these four expressions when the study was between subjects and included no previewing, no experimenter expectancy, and no forced-choice response format?

Available evidence suggests that the final result would vary with culture and with type of expression. For Western cultures, I do not believe that recognition for the "anger," "disgust," "fear," and "sad" expressions would decline to the level of chance, but evidence does not convince me that recognition would remain at 82%. The more important question is what would happen in non-Western societies. For non-Western literate cultures, recognition scores might hold for some expressions but fall lower for others. Overall, recognition might not

remain at 65%. And for cultures isolated from the West, such as the Fore, recognition scores might not remain at 49%. Recognition scores for some of these facial expressions might decline to the level of chance; recognition scores for others might fall to a low but nonrandom level. For isolated cultures, current evidence does not demonstrate reliable discrimination within the set of negative expressions when the threats to internal validity are removed.

A similar argument applies to the studies of freely generated labels. Although these studies eliminated some of the problems surrounding forced choice, they yielded lower initial recognition scores. Again, focus on "anger," "disgust," "fear," and "sadness" expressions. The median score (based on a liberal scoring scheme) for the hypothesized expressions of these four negative emotions was 61% for Western cultures, 54% for Japanese (based on only 1 sample), less than 30% for the Fore, and still less for the Bahimeno (see Table 18). From these scores, we subtract the influence of previewing, within-subject design, fixed order of presentation, experimenter expectancy, and other aids in conforming to the experimental hypothesis. Again, until more evidence is gathered, one can only guess what the final result would be. A baseline of chance cannot be as easily estimated here as in the case of forced choice. One baseline might be the base rate of generating a "correct" label when asked simply to name a negative emotion. In any case, for Western cultures, I do not believe that recognition scores would decline to such a baseline, but evidence does not convince me that recognition scores would remain at 61%. In non-Western cultures, the final recognition scores might be less high.

When recognition scores sink into the low or even moderate range, then new issues arise. If recognition scores drop below 50%, the possibility exists that the predicted interpretation is not the most common interpretation. (Such was the case in the Japanese free-label data: The modal response to the "fear" expression was surprise.) Wagner (1993) surveyed the way in which recognition was calculated and analyzed, finding that "almost all papers published between 1979 and 1992 . . . suffer from one or more of three errors" (p. 2): inappropriate measurement of recognition, the use in statistical analyses of inappropriate chance levels, and misapplication of chi-square and binomial statistical tests. When recognition scores are high, corrections for these errors are minor and do not alter the final conclusion. The lower the recognition scores, however, the greater the relative role of these errors in the results and the more likely the final conclusion will be altered.

Recognition scores for some facial expressions in some non-Western cultures might not exceed chance. Nevertheless, let us suppose that even when properly assessed and analyzed, recognition remains greater than chance for most facial expressions for all cultures. If so, the observer must provide some initial association between facial expressions and emotion labels. Method factors then could combine to guide that association into a consensual and clear pattern. Given the design of the standard experiment, even a vague beginning could result in a pattern that looked like the universality thesis. Note that on this account, method factors do not account for the results completely. In other words, if observers had no idea whatsoever

Table 18

Recognition Scores Estimated From Free-Label Responses for Four Negative Expressions

	Facial expression			
Culture	"Sadness" [a]	"Fear"	"Disgust" [b]	"Anger"
Western				
American[c]	60.1	59.6	49.0	66.4
Greek[c]	49.6	70.8	48.7	47.1
English[c]	63.5	58.9	46.4	56.6
French[c]	59.3	62.3	46.8	53.7
Canadian[d]	70	62	66	78
Greek[d]	75	87	68	63
Median	61.8	62.2	48.9	59.8
Literate non-Western				
Japanese[d]	80	14	56	48
Isolated non-Western				
Fore[e,f]	Negl	30	Negl	49
Bahimeno[e]	Negl	Negl	Negl	?

Note. Negl indicates that the figure was not reported by the original investigator but was apparently negligible.
[a] Izard's term was *distress.* [b] Combination of disgust and contempt in Izard's (1971) data. [c] Izard (1971); weighted average. [d] Russell, Suzuki, and Ishida (in press). [e] Sorenson (1976); see Table 16 and text for discussion. [f] Figure is median of three Fore groups.

about how to associate facial expressions with emotion labels, then the puzzle would remain unsolved. Method factors alone would not manufacture the associations. The standard method could help the subject think about questions that would not otherwise have arisen, and it might turn a vague idea into a highly consensual choice. But some association between face and emotion must be provided by the observer. So as great as the problems of method are, they cannot be used to dismiss the evidence entirely. Rather, alternative explanations of that evidence must be examined.

Alternative Interpretations

The studies reviewed here have been said to support specific theories of facial expression proposed by Ekman (1972, 1989) and Izard (1977). Strictly speaking, what these various studies did was rule out the null hypothesis of no agreement whatsoever about which emotions to attribute to various facial expressions. Ruling out random choice is not the same as confirming any particular theory. A study on the interpretation of facial expressions that pits the null hypothesis of random choice against any reasonable substantive hypothesis is likely to find in favor of that substantive hypothesis.

An explanation of the available evidence need not account for high consensus on the 7 ± 2 emotion categories. Rather, it need only provide an initial association between some facial configurations and some emotion labels. The initial association then provides a preliminary solution to the puzzle presented in the experiment. Method factors can then help shape this solution into the amount of agreement seen in the particular study. Various possibilities exist on just what this initial association might be, on how people spontaneously interpret facial movements. Here are several alternatives (these are alternatives to

each other rather than alternatives to the universality thesis— indeed some are versions of the universality thesis):

1. One possible explanation is an account along the lines proposed by Tomkins and McCarter (1964), Izard (1971, 1977), or Ekman and Friesen (1986). As I have attempted to demonstrate, the available evidence neither proves nor disproves these hypotheses. Their account might hold for some facial expressions, but not for others.

2. The observer might spontaneously interpret facial expressions in terms of a list of categories somewhat or completely different from those proposed in Alternative 1. For example, the list might consist of four broad categories—roughly, *happiness, surprise/fear, disgust/anger,* and *sadness.* This list might have the most cross-cultural viability. Or the list might be *calm, excited, surprised, distressed, frustrated, disgusted,* and *depressed.* Or the list might be much longer—studies of freely chosen labels suggest several hundred categories. On any of these alternatives, an observer would spontaneously interpret a particular facial expression in terms of a category often different from that supposed by the researcher. Given the standard forced-choice response format based on the researcher's hypothesis, the subject chooses the word on the list of options provided that is most similar. In this way, subjects' initial categorization would be channeled into the hypothesized categories.

The present alternative represents an uncountable number of specific hypotheses. The most interesting possibility is that the list of categories varies from one language to another, from one culture to another, or even from one person to another. Consider the Balinese, for whom "the face is taken to be the physical manifestation of hidden forces of major significance—the heart" (Wikan, 1990, p. 52). The categories into which the Balinese divide facial expressions are in some ways similar to but in some

ways different from what is done in English. Balinese distinguish *cerah muka* (clear, bright, happy) faces, *layu* (withered, sad) faces, *muram* (gloomy, cloudy, shamed) faces, *ny ebong* (grave, stern) faces, *sinis* (cynical) faces, and *galak* (jealous, furious) faces. More generally, the mental categories into which emotion is divided vary to some degree with language (Russell, 1991b; Wierzbicka, 1992).

3. The observer might initially interpret a facial expression in terms of bipolar dimensions rather than specific, discrete categories. Recall that when observers were asked to judge pleasure–displeasure from spontaneous facial expressions, they tended to be accurate. In a study such as Wagner et al.'s (1986), the results could be accounted for in terms of a single pleasant–unpleasant discrimination. More powerful still would be two- or three-dimensional accounts. Recall that Woodworth (1938) analyzed both agreements and disagreements and found that interpretation of facial expression could be understood as placement within broad clusters—which, in turn, fell along a single dimension. Schlosberg (1952, 1954) later proposed two and then three underlying dimensions. Subsequent researchers have found bipolar dimensions a useful way of accounting for aspects of how people conceptualize emotion (Daly, Lancee, & Polivy, 1983; Russell, 1980; Watson & Tellegen, 1985) and facial expressions (Abelson & Sermat, 1962; Russell & Bullock, 1986; Russell, Lewicka, & Niit, 1989). A review of cross-cultural evidence suggested that pleasantness–unpleasantness, arousal–sleep, and potency–submissiveness are candidates for universal dimensions in the conceptualization of emotion and feeling (Russell, 1991b). Young preschoolers categorize facial expressions in terms of the same broad dimensions (Bullock & Russell, 1986; Russell, 1989a).

To illustrate this type of account, imagine that subjects' initial interpretation of a face is in terms of degree of pleasure and degree of arousal. If the subject is given a forced-choice response format, then the subject chooses the option that comes closest in terms of pleasure and arousal. Even the simple two-dimensional account just described provides an explanation for various aspects of the data. Thus, consensus appears highest on *happiness* because the forced-choice format provides few or no alternative labels for faces judged as pleasant. For the negative emotion labels, 50% or 60% "correct" responses are easily understood. Suppose that an "anger" expression is perceived as conveying extreme displeasure and moderately high arousal. This combination occurs in labels such as *anger, fear,* and *disgust. Anger* is closest and chosen by the greatest number, *disgust* is next closest and is chosen by the second greatest number, and so on. This hypothesis also provides an account for the "confusions" reported by Ekman (1972) and others. For instance, the Dani of New Guinea "confused" (labeled similarly) "anger" and "disgust" expressions, which are similar in pleasure and arousal. The Fore of New Guinea confused "surprise" and "fear" expressions, which are similar in arousal.

The relation of a dimensional to a categorical account can be thought of in various ways, but one way concerns the amount of information inferred from a facial expression. Thus, in comparing a dimensional account to the categorical account implicit in many statements of the universality thesis, the number of categories recognized and the precise recognition scores become telling issues. Very high agreement on a large number of specific labels favors the categorical account, whereas lower agreement on fewer labels favors the dimensional account (the exact numbers required vary, of course, with the details of the judgment procedure). I have argued that the available evidence does not prove the categorical account even in Western college students. But even if the impact of method is less than I have estimated, the recognition scores in studies of non-Western societies are not so high as to force a categorical interpretation. Dimensions may therefore provide an account of what is universal in the interpretation of facial expression, whereas categories might be culture specific. For instance, one possibility is that the "anger" expression is universally (or widely) interpreted as conveying displeasure and moderately high arousal. In Western cultures, that expression is further interpreted as conveying frustration, determination, and a threat of violence. Displeasure, high arousal, frustration, determination, and threat of violence together suggest anger.

4. The observer might initially interpret a facial expression as a response to a type of situation. For instance, the observer might take a smile to be a response to the arrival of a friend, a wrinkled nose to be a response to a foul odor, raised eyebrows and wide open eyes to be a response to something new or unexpected, and an open mouth (and yelling) to be a response to danger. Seeing the facial expression allows an observer to infer the situation. Arrival of friend, foul odor, something new, and danger are types of situations, not internal states. Some of the labels seen in free-label studies, despite instructions to produce an emotion label, support this alternative: *frustration, bad news,* and *sees something pleasant.*

On this alternative, the situational interpretation is initial, spontaneous, and fast. An emotion might then be inferred, but more slowly and with more effort, and perhaps only if asked or only in some cultures. Thus, the observer quickly guesses that the person with a wrinkled nose is smelling something foul. Asked what emotion that person is feeling, the observer must work out an answer. If the experimenter allows a completely free response, the observer might say, "She's smelling a rotten egg" (Frijda, 1958). If the experimenter asks for a freely chosen emotion label, then different observers might reason to somewhat different answers: The person facing the foul smell might feel *nausea, upset, bewilderment, discomfort,* or some other emotion, depending on the specific circumstance imagined. If the experimenter provides a fixed list of emotions, the majority of observers might reason that of the options provided, disgust is the most appropriate answer, although a minority might reason otherwise.

5. The observer might initially interpret facial expression as part of an instrumental action. Indeed, there is a history of interpreting facial expressions in this way (Darwin, 1872/1965; Eibl-Eibesfeldt, 1973; Frijda, 1986). On such an account, for instance, the observer might take a smile to be an action with the purpose of greeting someone in a friendly (especially nonaggressive) manner. Wrinkling the nose has the purpose of blocking off an odor. Raising the eyebrows and widening the eyes have the purpose of seeing better. Lowering the brow while widening

the eyes has the purpose of threatening. Eibl-Eibesfeldt (1973) suggested that such facial actions were innate social signals— such as greetings, refusals, and threats—and that these same facial actions are innately responded to as such.

On this alternative, interpretation of a facial expression as a greeting, a refusal, or a threat is spontaneous, quick, automatic, and effortless. If asked, the subject in an experiment could, with effort, reason to a possible emotional state. Someone greeting a friend is more likely to be happy than sad, angry, lonely, and so on, so the reasoning might go. But the connection is not a necessary one: Someone can greet a friend with a smile and still be sad, lonely, and so on.

6. The observer might initially interpret not the whole facial expression, but its component actions. Scherer (1984), Smith (1989), and Ortony and Turner (1990) suggested that components of the facial expression occur separately and have separate meaning. For example, the hypothesized "facial expression of anger" consists of raised eyelids, wrinkled brow, compressed lips, and bared teeth. The raised eyelids are part of visual attention, the wrinkled brow a sign of frustration, the compression of lips a sign of resolve, and bared teeth part of an aggressive action. Interpretation of components would be quick, spontaneous, and effortless. Inference of an emotion would be based on interpretation of the components and would be slow, deliberate, and effortful—and perhaps culture-specific. Ekman (1992b) and Izard (1992) have commented on this account, and Turner and Ortony (1992) have replied.

7. Some of the hypotheses described above might be joined together into a fuller account. The dimensional account might be combined with expression-as-instrumental action. Thus an observer might simultaneously interpret a particular face as conveying displeasure and high arousal as well as a threat of violence. Alternatively, the meaning of a facial expression might be both situation and instrumental action. Fridlund (1991a) argued that current biological theory favors just such an account.

It would be possible to combine Alternatives 3, 5, and 6. According to one account along these lines (Fehr & Russell, 1984), each emotion word expresses a concept that is structured as a script. For instance, to attribute *anger* to someone is to see a resemblance between that person's current state and a generic script or prototypical case for *anger.* In the *anger* script, an offense occurs; the person responds with arousal and displeasure; an impulse to violence occurs and then results in action. Observing a facial movement, the observer automatically infers the situation, internal feelings, and the action. Then, with some additional effort, the observers use their cues to infer one or more scripts, which constitute emotion concepts. (For alternative uses of the notion of script, see Abelson, 1981, and Tomkins, 1984.)

8. The preceding hypotheses sometimes assume that the same hypothesis would apply to each of the "basic" facial expressions, to all cultures, and to all individuals within a culture. This assumption need not be correct. Regarding Alternative 2, for instance, it is also possible that the categories into which emotion is divided vary to some extent with language and culture (Russell, 1991b). If so, then the exact message inferred from a facial expression might vary (within limits) with culture.

What actually occurs might involve some complex mixture of the above alternatives, a mixture that varies from one culture, or even one individual, to another. In one culture, or for one individual, a smile might be interpreted as indicating happiness (an emotion), a wrinkled nose as indicating a foul odor (situation), and lowered brows with staring eyes as indicating a threat (action). In another culture, or for another individual, some other mixture might apply. Another possibility is that some cultures or individuals have no habitual way to interpret a specific facial expression but must create an interpretation on the spot.

Conclusion. We Westerners might find some of these alternative explanations implausible, but to assume that all people in all cultures think as we do would be to assume what the studies under review were designed to test. The available evidence is consistent with any of the preceding eight hypotheses—and undoubtedly more: These eight simply illustrate the existence of alternative explanations. The only alternative I mean to leave out is the null hypothesis of completely random association between a face and the emotion attributed to it. Some association is assumed, and the question is the nature of the association and the role of culture in it. The eight alternative explanations listed here, although obviously preliminary, may provide a beginning in answering that question. The eight are not necessarily mutually exclusive, and may all be part of the truth. Such a thought is comforting, but a more useful tactic in planning research is to view these as competing hypotheses. As a minimum, future research can include comparison levels that are more telling than random assortment. For example, I used pleasantness–unpleasantness as a comparison level in analyzing several studies in this review. Even moving from a one-dimensional (pleasant–unpleasant) to a two-dimensional baseline would be informative. More interesting comparison levels would be created by more complex accounts.

Conclusion

This article did not address the adequacies of different theories of emotion; the existence, discreteness, basicness, or universality of specific emotions; the origin of facial behavior; genetic and epigenetic influences on facial behavior; or other related topics. Nor is this article the final word on the topic it did address. Articles in psychology exist in a social dialogue. My turn was aimed at raising questions about a small but influential set of studies. Relevant evidence might have escaped my selection criteria. In their turn, proponents of alternative views may cite evidence that answers the questions raised here. New researchers will bring forward new evidence and new methods.

This is a topic on which opinions can differ. The merits of alternative explanations cannot now be decided on the basis of the empirical evidence available and are therefore decided on the basis of a subjective judgment of plausibility. With concepts such as emotion, anger, fear, and so on lacking a clear consensual definition and with the universality thesis a background assumption of our everyday thought, many readers of this literature might find the universality thesis the most plausible alternative available. On the other hand, those who wait until the evidence compels them to decide must seek further evidence.

I hope that this article provides an opportunity for a renewal of the basic questions that initiated this field of study. We have the advantage today of new techniques for the measurement of facial behavior (Cacioppo & Petty, 1983; Ekman & Friesen, 1982; Izard, 1983). We have new ideas on the processes involved in the production of facial behavior (Fridlund, 1991a; Ortony & Turner, 1990; Scherer, 1984; Zajonc, Murphy, & Inglehart, 1989). And we have hypotheses to refine and test about the meaning that observers attribute to facial behavior, such as the eight hypotheses just listed.

How might future research best be approached? As a first step, we need to abandon any implicit assumption that we have only two alternatives: randomness and universality. This forced choice underestimates our options. Ambiguities in phrases such as "universality of facial expression" must then be reduced. More precise statements of the universality thesis, and indeed of all the alternatives listed above, would aid in the design of more useful studies. The assumptions and implications for each alternative must be spelled out. How much recognition does it predict? In what contexts does it apply? For each hypothesis, it is helpful to specify what it predicts will not happen and which alternatives it excludes. A version of the universality thesis compatible with all or most of the eight alternatives listed above is too vague to remain scientifically useful. As long as a hypothesis remains vague or ambiguous, the results of empirical efforts will be difficult to assess.

In gathering new data, the issues of internal, convergent, and ecological validity must be raised earlier and more often. Doing so would encourage the development and use of new and multiple methods rather than reliance on any one. In raising the issue of ecological validity, I am not advocating an abandonment of experimental techniques. Rather, the method chosen must suit the question asked, and I believe that ecological questions should be among those asked. Even when experimental techniques are used, ecologically oriented questions should be borne in mind: How often do the stimuli being used occur, in that strength, for that length of time, in natural settings? How reflective is the response measure of naturally occurring interpretations given to the facial expressions observed? Recall how little information is available on the spontaneous interpretations given to spontaneous facial actions, especially in unfamiliar cultures. We need basic descriptive information on the sorts of facial movements that occur in everyday settings, with what frequency and in what contexts. When facial movements occur, we need to know how observers naturally interpret them, with what frequency and in what contexts. A breakthrough in the study of facial expression in infants came from looking at the actual behavior of one infant in actual situations (Camras, 1992).

Facial movements include not just the 7 ± 2 "facial expressions of emotion," but laughs, pouts, yawns, winces, grimaces, and all manner of actions difficult to describe. In our culture, people use facial cues to infer sleepiness, relaxation, puzzlement, confusion, pain, boredom, interest, attention, and other states besides seven "basic" emotions. An understanding of facial expressions would be helped by integrating studies over the full range of facial movements and inferences from them.

Understanding facial expression requires more than just a change in method but also changes in the assumptions underlying the choice of method. To illustrate, consider this question: Why did the method I have called *standard* come to be used more than any other? I have argued so far that methodological rigor and ecological validity were not the reasons for choosing the standard method. The standard method evolved and survived. What was the mechanism of selection? Since initial research in the 19th century, various methods have been used, but most were abandoned. Spontaneous facial expressions were studied (Landis, 1924; Munn, 1940; Sherman, 1927; Vinacke, 1949), but posed expressions worked better. Subjects were allowed to describe their interpretation of the facial expression in their own words (Darwin, 1872/1965; Frijda, 1953), but forced choice worked better. To *work* is to produce plausible or meaningful results. As one colleague put it, the standard method produced "positive results." Plausible, meaningful, or positive results are those that make most sense to someone with our cultural background. In short, the mechanism underlying the evolution of the standard method is something that has existed since the time of Aristotle, the commonsense appeal of the universality thesis.

Tellingly, even the standard method was abandoned when it stopped working. In the first studies of isolated cultures, Ekman et al. (1969) used the standard method, but it worked only moderately well. They then turned to other methods, principally the Dashiell technique, in which a story was substituted for the emotion name. And this new method worked much better.

A similar process was also implicit in reviews of the relevant literature. Studies less supportive of universality, mainly carried out in the 1920s, are often forgotten. Studies reported in or after 1969 are more emphasized, the more their results agree with universality. Results within a study such as lower or nonsignificant agreement with prediction for particular facial expressions are downplayed. (Examine the conclusions in Table 1.) Nonsupportive studies, anomalies in the data, methodological problems, and other qualifications gradually fade from sight. As the evidence makes its way into a review of the literature or a textbook, and then is cited in articles or chapters, that evidence tends to be gradually assimilated to the background assumption, the universality thesis.

The universality thesis is an idea that we Western psychologists find plausible, especially given randomness as the alternative. We speakers of English find it plausible that our concepts of *anger, fear, contempt,* and the like are universal categories, exposing nature at the joints. One way to overcome the influence of such implicit assumptions is to emphasize alternative conceptualizations. And, I believe, the most interesting means to this end is to take seriously the conceptualizations (ethnotheories, cultural models) found in other cultures. Rather than ask whether a given culture agrees with one preformulated hypothesis, we might more usefully ask how members of that culture conceptualize emotions and facial behavior. There may be no short cut to obtaining the needed information. Although the task is great, what we know about the peoples of different cultures suggests that carrying it out will be fascinating. In studying the beliefs of other cultures, we cannot expect to know right

away whether those beliefs are correct. Their theories are not to be believed any more than our own. I doubt whether even asking about their correctness is useful. We don't yet know whether Aristotle was right in his beliefs regarding physiognomy: whether soft hair reveals a coward or whether facial expressions reveal emotions. We might more usefully gather the beliefs of different cultures rather than evaluate them.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36*, 715–729.

Abelson, R. P., & Sermat, V. (1962). Multidimensional scaling of facial expressions. *Journal of Experimental Psychology, 63*, 546–554.

Allport, F. H. (1924). *Social psychology.* Boston: Houghton Mifflin.

Aristotle. (1913). Physiognomonica. In W. D. Ross (Ed.) and T. Loveday & E. S. Forster (Trans.), *The works of Aristotle* (pp. 805–813). Oxford, England: Clarendon.

Bain, A. (1855). *The senses and the intellect.* London: Parker.

Bain, A. (1859). *The emotions and the will.* London: Parker.

Bell, C. (1806). *Essays on the anatomy of expression in painting.* London: Longman, Hurst, Rees, & Orme.

Birdwhistell, R. L. (1970). *Kinesics and context.* Philadelphia: University of Pennsylvania Press.

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

Braun, J., & Linder, D. E. (1979). *Psychology today* (4th ed.). New York: Random House.

Brown, D. E. (1991). *Human universals.* Philadelphia: Temple University Press.

Brown, R. (1986). *Social psychology* (2nd ed.). New York: Free Press.

Bruner, J. S., & Tagiuiri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2). Cambridge: Addison-Wesley.

Buck, R. (1988). *Human motivation and emotion* (2nd ed.). New York: Wiley.

Buck, R. W, Miller, R. E., & Caul, W. F. (1974). Sex, personality, and physiological variables in the communication of emotion via facial expression. *Journal of Personality and Social Psychology, 30*, 587–596.

Buck, R. W., Savin, V. J., Miller, R. E., & Caul, W. F. (1972). Communication of affect through facial expressions in humans. *Journal of Personality and Social Psychology, 23*, 362–371.

Bullock, M., & Russell, J. A. (1986). Concepts of emotion in developmental psychology. In C. E. Izard & P. B. Read (Eds.), *Measuring emotions in infants and children* (Vol. 2). Cambridge, England: Cambridge University Press.

Buss, D. (1992). Is there a universal human nature? *Contemporary Psychology, 37*, 1262–1263.

Buzby, D. E. (1924). The interpretation of facial expression. *American Journal of Psychology, 35*, 602–604.

Cacioppo, J. T., & Petty, R. E. (1983). *Social psychophysiology: A sourcebook.* New York: Guilford Press.

Camras, L. (1992). A dynamic systems perspective on expressive development. In K. Strongman (Ed.), *International review of studies on emotion* (pp. 16–28). New York: Wiley.

Carlson, J. G., & Hatfield, E. (1992). *Psychology of emotion.* Fort Worth, TX: Harcourt Brace Jovanovich.

Chan, D. W. (1985). Perception and judgment of facial expressions among the Chinese. *International Journal of Psychology, 20*, 681–692.

Cüceloglu, D. M. (1970). Perception of facial expressions in three cultures. *Ergonomics, 13*, 93–100.

Daly, E. M., Lancee, W., & Polivy, J. (1983). A conical model for the taxonomy of emotional experience. *Journal of Personality and Social Psychology, 45*, 443–457.

Darwin, C. (1965). *The expression of the emotions in man and animals.* Chicago: University of Chicago Press. (Original work published 1872)

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin, 111*, 203–243.

Ducci, L., Arcuri, L., W/Georgis, T., & Sineshaw, T. (1982). Emotion recognition in Ethiopia. *Journal of Cross-Cultural Psychology, 13*, 340–351.

Duchenne de Boulogne, G. B. (1990). *The mechanism of human facial expression* (R. A. Cuthbertson, Ed. and Trans.). Cambridge, England: Cambridge University Press. (Original work published 1862)

Eibl-Eibesfeldt, I. (1972). Similiarities and differences between cultures in expressive movements. In R. A. Hinde (Ed.), *Nonverbal communication* (pp. 297–311). Cambridge, England: Cambridge University Press.

Eibl-Eibesfeldt, I. (1973). The expressive behavior of the deaf-and-blind-born. In M. von Cranach & I. Vine (Eds.), *Social communication and movement* (pp. 163–194). San Diego, CA: Academic Press.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation, 1971* (pp. 207–283). Lincoln: University of Nebraska Press.

Ekman, P. (1973). *Darwin and facial expression: A century of research in review.* San Diego, CA: Academic Press.

Ekman, P. (1980). *The face of man: Expressions of universal emotions in a New Guinea village.* New York: Garland STPM Press.

Ekman, P. (Ed.). (1982). *Emotion in the human face* (2nd ed.). Cambridge, England: University of Cambridge Press.

Ekman, P. (1984). Expression and the nature of emotion. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Hillsdale, NJ: Erlbaum.

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

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

Ekman, P. (1992b). Are there basic emotions? *Psychological Review, 99*, 550–553.

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

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

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face.* Englewood Cliffs, NJ: Prentice Hall.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Measuring facial movement with the facial action coding system. In P. Ekman (Ed.), *Emotion in the human face* (2nd ed., pp. 178–211). Cambridge, England: Cambridge University Press.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion, 10*, 159–168.

Ekman, P., & Friesen, W. V. (1988). Who knows what about contempt: A reply to Izard and Haynes. *Motivation and Emotion, 12*, 17–22.

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

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face.* New York: Pergamon Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tar-

latzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.

Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion, 12*, 303–308.

Ekman, P., Heider, E., Friesen, W. V., & Heider, K. (1972). Unpublished raw data.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology, 30*, 527–554.

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991a). Confusions about context in the judgment of facial expression: A reply to "Contempt and the relativity thesis." *Motivation and Emotion, 15*, 169–176.

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991b). Contradictions in the study of contempt: What's it all about? Reply to Russell. *Motivation and Emotion, 15*, 293–296.

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

Evans, E. C. (1969). Physiognomics in the ancient world. *Transactions of the American Philosophical Society, 59*, 1–101.

Faigin, G. (1990). *The artist's complete guide to facial expression.* New York: Watson-Guptill.

Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113*, 464–486.

Feleky, A. M. (1914). The expression of the emotions. *Psychological Review, 21*, 33–41.

Feleky, A. M. (1922). *Feelings and emotions.* New York: Pioneer Press.

Fernandez-Dols, J-M. (in press). Il ruolo del contesto nell'attribuzione del contenuto emozionale alle espressioni facciali [The role of the context in the attribution of emotional content to facial expressions]. In G. Bellelli (Ed.), *Sapere e sentire* [Knowing and feeling]. Napoli, Italy: Liguori.

Fernandez-Dols, J.-M., Sierra, B., & Ruiz-Belda, M. A. (1993). On the clarity of expressive and contextual information in the recognition of emotions: A methodological critique. *European Journal of Social Psychology, 23*, 195–202.

Fernandez-Dols, J.-M., Wallbott, H., & Sanchez, F. (1991). Emotion category accessibility and the decoding of emotion from facial expression and context. *Journal of Nonverbal Behavior, 15*, 107–123.

Fernberger, S. W. (1928). False suggestion and the Piderit model. *American Journal of Psychology, 40*, 562–568.

Fridlund, A. J. (1991a). Evolution and facial action in reflex, social motive, and paralanguage. *Biological Psychology, 32*, 3–100.

Fridlund, A. J. (1991b). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60*, 229–240.

Fridlund, A. J. (1992). Darwin's anti-Darwinism in "The expression of the emotions in man and animals." In K. Strongman (Ed.), *International review of emotion* (Vol. 2, pp. 117–138). New York: Wiley.

Fridlund, A. J., Ekman, P., & Oster, H. (1987). Facial expressions of emotion: Review of the literature, 1970–1983. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication* (pp. 143–224). Hillsdale, NJ: Erlbaum.

Frijda, N. H. (1953). The understanding of facial expression of emotion. *Acta Psychologica, 9*, 294–362.

Frijda, N. H. (1958). Facial expression and situational cues. *Journal of Abnormal and Social Psychology, 57*, 149–154.

Frijda, N. H. (1986). *The emotions.* Cambridge, England: Cambridge University Press.

Goldberg, H. D. (1951). The role of "cutting" in the perception of the motion picture. *Journal of Applied Psychology, 35*, 70–71.

Goodenough, F. L. (1931). The expression of the emotions in infancy. *Child Development, 2*, 96–101.

Goodenough, F. L. (1932). The expression of the emotions in a blind-deaf child. *Journal of Abnormal Psychology, 27*, 328–333.

Goodenough, F. L., & Tinker, M. A. (1931). The relative potency of facial expression and verbal description of stimulus in the judgment of emotion. *Journal of Comparative Psychology, 12*, 365–370.

Gudykunst, W. B., & Ting-Toomey, S. (1988). *Culture and interpersonal communication.* Newbury Park, CA: Sage.

Guilford, J. P. (1929). An experiment in learning to read facial expression. *Journal of Comparative Psychology, 24*, 191–202.

Helson, H. (1964). *Adaptation-level theory.* New York: Harper & Row.

Howell, R. J., & Jorgenson, E. (1970). Accuracy of judging emotional behavior in a natural setting: A replication. *Journal of Social Psychology, 81*, 269–270.

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

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

Izard, C. E. (1980). Cross-cultural perspectives on emotion and emotion communication. In H. Triandis & W. Lonner (Eds.), *Handbook of cross-cultural psychology: Basic processes* (Vol. 3, pp. 185–222). Boston: Allyn & Bacon.

Izard, C. E. (1983). *The maximally discriminative facial movement coding system* (Rev. ed.). Newark, DE: Instructional Resources Center.

Izard, C. E. (1990). Facial expressions and the regulation of emotions. *Journal of Personality and Social Psychology, 58*, 487–498.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion–cognition relations. *Psychological Review, 99*, 561–565.

Izard, C. E., & Haynes, O. M. (1988). On the form and universality of the contempt expression: A challenge to Ekman and Friesen's claim of discovery. *Motivation and Emotion, 12*, 1–16.

Izard, C. E., & Malatesta, C. Z. (1987). Perspectives on emotional development. I: Differential emotions theory of early emotional development. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 494–554). New York: Wiley.

Izard, C. E., & Saxton, P. M. (1988). Emotions. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., Vol. 1, pp. 627–676). New York: Wiley.

Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion, 3*, 81–123.

Kagan, J., & Havemann, E. (1976). *Psychology: An introduction.* New York: Harcourt Brace Jovanovich.

Kilbride, J. E., & Yarczower, M. (1976). Recognition of happy and sad facial expressions among Baganda and U.S. children. *Journal of Cross-Cultural Psychology, 7*, 181–194.

Kilbride, J. E., & Yarczower, M. (1980). Recognition and imitation of facial expressions: A cross-cultural comparison between Zambia and the United States. *Journal of Cross-Cultural Psychology, 11*, 281–296.

Kilbride, J. E., & Yarczower, M. (1983). Ethnic bias in the recognition of facial expressions. *Journal of Nonverbal Behavior, 8*, 27–41.

Klineberg, O. (1938). Emotional expression in Chinese literature. *Journal of Abnormal and Social Psychology, 33*, 517–520.

Klineberg, O. (1940). *Social psychology.* New York: Holt.

Knudsen, H. R., & Muzekari, L. H. (1983). The effects of verbal statements of context on facial expressions of emotion. *Journal of Nonverbal Behavior, 7*, 202–212.

LaBarre, W. (1947). The cultural basis of emotions and gestures. *Journal of Personality, 16*, 49–68.

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

Langfeld, H. S. (1918a). The judgment of emotions from facial expressions. *Journal of Abnormal and Social Psychology, 13*, 172–184.

Langfeld, H. S. (1918b). Judgments of facial expression and suggestion. *Psychological Review, 25*, 488–494.

Leach, E. (1972). The influence of cultural context on nonverbal communication in man. In R. Hinde (Ed.), *Nonverbal communication* (pp. 212–227). Cambridge, England: Cambridge University Press.

Le Brun, C. (1982). *Méthode pour apprendre à dessiner les passions, proposée dans une conference sur l'expression generale et particuliere* [Method for learning to draw the emotions, proposed in a lecture on particular expressions and expression in general]. Hildesheim, West Germany: Verlag. (Original work published 1702)

London, P. (1978). *Beginning psychology* (Rev. ed.). Homewood, IL: Dorsey.

Lorenz, K. (1965). *Evolution and modification of behavior.* Chicago: University of Chicago Press.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging, 2*, 64–69.

Manis, M. (1967). Context effects in communication. *Journal of Personality and Social Psychology, 5*, 326–334.

Manis, M. (1971). Context effects in communication. In M. H. Appley (Ed.), *Adaptation level theory* (pp. 237–255). San Diego, CA: Academic Press.

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

Matsumoto, D. (1992a). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology, 23*, 72–84.

Matsumoto, D. (1992b). More evidence for the universality of a contempt expression. *Motivation and Emotion, 16*, 363–368.

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE).* Unpublished slide set and brochure, San Francisco State University, Department of Psychology.

Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion, 13*, 143–157.

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expression of emotion. *Journal of Cross-Cultural Psychology, 17*, 211–224.

Mead, M. (1975). Review of "Darwin and facial expression." *Journal of Communication, 25*, 209–213.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112*, 179–204.

Montgomery, W. (1985). Charles Darwin's thought on expressive mechanisms in evolution. In G. Zivin (Ed.), *The development of expressive behavior* (pp. 27–50). San Diego, CA: Academic Press.

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication, 52*, 1–22.

Munn, N. L. (1940). The effect of knowledge of the situation upon judgment of emotion from facial expressions. *Journal of Abnormal and Social Psychology, 35*, 324–338.

Nakamura, M., Buck, R., & Kenny, D. A. (1990). Relative contribution of expressive behavior and contextual information to the judgment of the emotional state of another. *Journal of Personality and Social Psychology, 59*, 1032–1039.

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: Mechanisms of development. *Child Development, 58*, 889–909.

Niit, T., & Valsiner, J. (1977). Recognition of facial expressions: An experimental investigation of Ekman's model. *Acta et Commentationes Universitatis Tarvensis, 429*, 85–107.

Oatley, K. (1992). *Best laid schemes: The psychology of emotions.* Cambridge, England: Cambridge University Press.

Oatley, K., & Jenkins, J. M. (1992). Human emotions: Function and dysfunction. *Annual Review of Psychology, 43*, 55–85.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 74*, 315–341.

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology, 7*, 1–30.

Oster, H., Daily, L., & Goldenthal, P. (1989). Processing facial affect. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 107–161). Amsterdam: Elsevier.

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

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72*, 407–418.

Peck, S. R. (1987). *Atlas of facial expression.* New York: Oxford University Press.

Piderit, T. (1867). *Wissenschaftliches system der mimik und physiognomik* [Scientific system of mimicry and physiognomy]. Detmold, Germany: Kingenberg.

Reuter-Lorenz, P., & Davidson, R. J. (1981). Differential contributions of the two cerebral hemispheres to the perception of happy and sad faces. *Neuropsychologia, 19*, 609–613.

Ricci-Bitti, P. E., Brighetti, G., Garotti, P. L., & Boggi-Cavallo, P. (1989). Is contempt expressed by pancultural facial movements? *Recent Advances in Social Psychology: An International Perspective*, 329–339.

Rinn, W. E. (1984). The neuropsychology of facial expressions: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95*, 52–77.

Robarchek, C. (1977). Frustration, aggression, and the nonviolent Semai. *American Ethnologist, 4*, 762–779.

Ruckmick, C. A. (1921). A preliminary study of the emotions. *Psychological Monographs, 30*(3, Whole No. 136).

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

Russell, J. A. (1989a). Culture, scripts, and children's understanding of emotion. In C. Saarni & P. L. Harris (Eds.), *Children's understanding of emotion* (pp. 293–318). Cambridge, England: Cambridge University Press.

Russell, J. A. (1989b). Measures of emotion. In R. Plutchik & H. Kellerman (Ed.), *Emotion: Theory, research, and experience* (Vol. 4, pp. 83–112). San Diego, CA: Academic Press.

Russell, J. A. (1991a). The contempt expression and the relativity thesis. *Motivation and Emotion, 15*, 149–168.

Russell, J. A. (1991b). Culture and the categorization of emotion. *Psychological Bulletin, 110*, 426–450.

Russell, J. A. (1991c). Negative results on a reported facial expression of contempt. *Motivation and Emotion, 15*, 281–291.

Russell, J. A. (1993a). Forced-choice response format in the study of facial expression. *Motivation and Emotion, 17*, 41–51.

Russell, J. A. (1993b). *Stimulus presentation in the study of facial expression.* Manuscript submitted for publication.

Russell, J. A., & Bullock, M. (1986). Fuzzy concepts and the perception of emotion in facial expression. *Social Cognition, 4*, 309–341.

Russell, J. A., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General, 116,* 223–237.

Russell, J. A., & Lanius, U. F. (1984). Adaptation level and the affective appraisal of environments. *Journal of Environmental Psychology, 4,* 119–135.

Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology, 57,* 848–856.

Russell, J. A., Suzuki, N., & Ishida, N. (in press). Freely produced labels for facial expressions of emotion. *Motivation and Emotion.*

Scherer, K. (1984). On the nature and function of emotions: A component process approach. In K. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.

Scherer, K. (1992). What does a facial expression express? In K. Strongman (Ed.), *International review of studies on emotion* (pp. 139–165). New York: Wiley.

Schlosberg, H. (1941). A scale for judgment of facial expressions. *Journal of Experimental Psychology, 29,* 497–510.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology, 44,* 229–237.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61,* 81–88.

Schulze, R. (1912). *Experimental psychology and pedagogy.* New York: Macmillan.

Sherman, M. (1927). The differentiation of emotional responses in infants: II. The ability of observers to judge the emotional characteristics of the crying of infants and of the voice of an adult. *Journal of Comparative Psychology, 7,* 335–351.

Shimoda, K., Argyle, M., & Ricci-Bitti, P. (1978). The intercultural recognition of emotional expressions by three national racial groups: English, Italian and Japanese. *European Journal of Social Psychology, 8,* 169–179.

Skinner, M., & Mullen, B. (1991). Facial asymmetry in emotional expression: A meta-analysis of research. *British Journal of Social Psychology, 30,* 113–124.

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology, 56,* 339–353.

Sorenson, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (Eds.), *Psychological anthropology* (pp. 361–372). Chicago: Aldine.

Sorenson, E. R. (1976). *The edge of the forest: Land, childhood and change in a New Guinea protoagricultural society.* Washington, DC: Smithsonian Institution Press.

Spencer, H. (1855). *The principles of psychology.* Longman, Brown, Green & Longman: London.

Spignesi, A., & Shor, R. (1981). The judgment of emotion from facial expressions, contexts and their combination. *Journal of General Psychology, 104,* 41–58.

Stratton, G. M. (1921). The control of another person by obscure signs. *Psychological Review, 28,* 301–314.

Tanaka-Matsumi, J., Nelson, S., Attivissimo, D., & D'Urso, T. (1993). *Positive and negative context effects on the judgment of basic emotions in the face.* Manuscript submitted for publication.

Thayer, S. (1980a). The effect of expression sequence and expressor identity on judgments of the intensity of facial expression. *Journal of Nonverbal Behaivor, 5,* 71–79.

Thayer, S. (1980b). The effect of facial expression sequence upon judgments of emotion. *Journal of Social Psychology, 111,* 305–306.

Thompson, J. (1941). Development of facial expression of emotion in blind and seeing children. *Archives of Psychology, 37,* 1–47.

Tomkins, S. S. (1962–1963). *Affect, imagery, consciousness.* (Volumes 1 and 2). New York: Springer.

Tomkins, S. S. (1980). Affect as amplification: Some modifications in theory. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (Vol. 1, pp. 141–187). San Diego, CA: Academic Press.

Tomkins, S. S. (1984). Affect theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 163–195). Hillsdale, NJ: Erlbaum.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, 18,* 119–158.

Triandis, H. C., & Lambert, W. W. (1958). A restatement and test of Schlosberg's theory of emotion with two kinds of subjects from Greece. *Journal of Abnormal and Social Psychology, 56,* 321–328.

Turner, T. J., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review, 99,* 566–571.

Vinacke, W. E. (1949). The judgment of facial expressions by three national-racial groups in Hawaii: I. Caucasian faces. *Journal of Personality, 17,* 407–429.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 1–28.

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

Wallbott, H. G. (1988). In and out of context: Influences of facial expression and context information on emotion attributions. *British Journal of Social Psychology, 27,* 357–369.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.

Watson, S. G. (1972). Judgment of emotion from facial and contextual cue combinations. *Journal of Personality and Social Psychology, 24,* 334–342.

Wierzbicka, A. (1992). *Semantics, culture and cognition.* New York: Oxford University Press.

Wikan, U. (1990). *Managing turbulent hearts: A Balinese formula for living.* Chicago: University of Chicago Press.

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

Wolfgang, A., & Cohen, M. (1988). Sensitivity of Canadians, Latin Americans, Ethiopians, and Israelis to interracial facial expressions of emotions. *International Journal of Intercultural Relations, 12,* 139–151.

Woodworth, R. S. (1938). *Experimental psychology.* New York: Holt.

Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology.* New York: Holt.

Wyer, R. S., Jr., & Srull, T. K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 161–197). Hillsdale, NJ: Erlbaum.

Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: Implications of the vascular theory of emotion. *Psychological Review, 96,* 395–416.