

# On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis

Hillary Anger Elfenbein and Nalini Ambady  
Harvard University

A meta-analysis examined emotion recognition within and across cultures. Emotions were universally recognized at better-than-chance levels. Accuracy was higher when emotions were both expressed and recognized by members of the same national, ethnic, or regional group, suggesting an in-group advantage. This advantage was smaller for cultural groups with greater exposure to one another, measured in terms of living in the same nation, physical proximity, and telephone communication. Majority group members were poorer at judging minority group members than the reverse. Cross-cultural accuracy was lower in studies that used a balanced research design, and higher in studies that used imitation rather than posed or spontaneous emotional expressions. Attributes of study design appeared not to moderate the size of the in-group advantage.

Psychologists have long debated whether emotions are universal versus whether they vary by culture. These issues have been extensively summarized elsewhere and we do not reiterate them (e.g., Ekman, 1972, 1994; Izard, 1971; Mesquita & Frijda, 1992; Mesquita, Frijda, & Scherer, 1997; Russell, 1994; Scherer & Wallbott, 1994). Although many theorists have taken extreme positions and provoked lively debate, recent theoretical models have attempted to account for both universality and cultural variation by specifying which particular aspects of emotion show similarities and differences across cultural boundaries (e.g., Ekman, 1972; Fiske, Kitayama, Markus, & Nisbett, 1998; Markus &

Kitayama, 1991; Mesquita & Frijda, 1992; Mesquita et al., 1997; Scherer, 1997; Scherer & Wallbott, 1994). Many of these models present an interactionist perspective (Ekman, 1994; Matsumoto, 1989; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Russell, 1994; Scherer & Wallbott, 1994). For example, Matsumoto (1989) argued that, although emotions are biologically programmed, the process of learning to control both their expression and perception is highly dependent on cultural factors. Taking a different perspective, Russell (1994) suggested that specific emotional categories are largely culturally specific but that broad dimensions such as valence and arousal are universal.

Although these models are rich and illuminating, Scherer and Wallbott (1994) have noted that there have not been many systematic attempts to gather evidence to examine the theories. Attempting to garner such evidence, Scherer and Wallbott found strong evidence for universality as well as cultural differences in emotional experience, including both psychological and physiological responses to emotions. The primary goal of the present article is to conduct a meta-analysis to examine the evidence that exists for both universality and cultural differences in the particular area of *emotion recognition*. Emotional expression falls under the category of *emotional behavior*: the outward expressions and actions that accompany emotional experience (Mesquita & Frijda, 1992; Mesquita et al., 1997). Because such emotional behaviors are observable, they have been relatively more available for study by researchers who have examined both the expressions themselves as well as their accurate recognition. As a consequence, there is a rich literature on cross-cultural emotion recognition that can be empirically examined in order to assess the universality and the cultural specificity of emotion recognition, as well as to identify the factors that moderate these phenomena.

---

Hillary Anger Elfenbein, Program in Organizational Behavior, Harvard University; Nalini Ambady, Department of Psychology, Harvard University.

Preparation of this article was supported by a National Science Foundation Graduate Student Fellowship and Presidential Early Career Award for Scientists and Engineers Grant BCS-9733706.

We thank David Matsumoto and Klaus Scherer for their extremely helpful comments. We are grateful to each author who responded to our requests for articles from their file drawers. We would also like to express appreciation to Andrew Calder, David Chan, Pierre Gosselin, William Ickes, Gilles Kirouac, Manas Mandal, Roslyn Markham, David Matsumoto, Steve Nowicki, Pio Ricci Bitti, Klaus Scherer, and H. L. Toner, who provided not only manuscripts but also reached deep into their archives to provide us with detailed information about their studies. We thank Donald Rubin, Robert Rosenthal, and the late Douwe Yntema for suggestions about the statistical analysis. For helpful comments on an earlier draft, we thank Teresa Amabile, Debi La Plante, Abby Marsh, Todd Pittinsky, Jennifer Richeson, and Margaret Shih. For assistance with translations, we thank Oksana Cherniavska (Russian), Daniel Elfenbein (German), Ying Liu (Chinese), Paolo Martini (Italian), and Eiichi Miyasaka (Japanese). For research assistance, we thank Heather Gray. We also thank the tireless staff of the interlibrary loan desk at Widener Library.

Correspondence concerning this article should be addressed to Hillary Anger Elfenbein, Harvard Business School, Baker Library, Soldiers Field, Boston, Massachusetts 02163. E-mail: hillary@post.harvard.edu. A supplementary report including additional detail on analyses is available upon request from Hillary Anger Elfenbein.

## Universality and Cultural Specificity in Emotion Recognition

Proponents of emotional universality have used cross-cultural studies of facial emotion recognition as one of their central sources of evidence. Classic studies conducted by Paul Ekman (1972),

Carroll Izard (1971), and their colleagues have demonstrated that facial photographs of Americans expressing “basic” emotions can be recognized at above-chance accuracy in both literate and preliterate cultures, implying that the six (Ekman, 1972) or seven (Biehl et al., 1997) facial expressions of emotions used in their experiments are universally recognized.

There has been some criticism of these emotion recognition studies as one of several sources of evidence for emotional universality. First, as we discuss below, the results of such studies are smaller with certain samples, such as members of preliterate groups. Second, as Mesquita and Frijda (1992) have noted, the stimuli may lack ecological validity. Studies using spontaneous facial expressions have attempted to address this validity issue directly (e.g., Ekman & Rosenberg, 1997). Finally, emotion recognition studies demonstrating universality have been criticized on methodological grounds. For example, Russell (1994) has argued that studies indicating universality understate cultural differences through the use of methodological artifacts such as forced-choice response formats and posed expressions. However, these arguments themselves have been the subject of debate (e.g., Ekman, 1994; Izard, 1994).

The interest of the original researchers conducting the classic studies of emotion recognition was to investigate the universality of emotion recognition. Thus, in most of the early studies, cultural differences in the original cross-cultural data were not examined “because the researchers were interested at that time in exploring agreement, not disagreement” (Matsumoto & Assar, 1992, p. 86). But the same studies that provided evidence for universality also hinted at cultural specificity. For example, Ekman and his colleagues (Ekman, Sorensen, & Friesen, 1969; Sorensen, 1975) tested members of several tribes in New Guinea that had almost no contact with Westerners. Although their results indicated that preliterate participants were capable of accuracy at above-chance levels in certain experimental conditions, accuracy was still less than that achieved by literate participants. In the most extreme case, the Bahinemo tribes people saw all the faces of Americans as angry, and they had “no clear mode of response to specific pictures” (Sorensen, 1975, p. 367).

More recent attention has been directed toward examining cultural differences in these same data (e.g., Matsumoto, 1989; Mesquita & Frijda, 1992; Mesquita et al., 1997; Russell, 1994). Although accuracy levels for most groups in the studies were significantly above the levels expected from chance guessing, European and American participants generally scored higher in these studies using American expressors than did Asian or African participants (e.g., Ekman, 1972; Izard, 1971). For example, in Izard’s (1971) large-scale study, American and European groups correctly identified 75–83% of the facial photographs, whereas Japanese groups identified 65% and African groups correctly identified only 50%. Izard (1971) omitted these African data from further analysis because the particular participants did not complete the task in their native language, although other African samples have shown similar results. In a study by Ducci, Arcuri, W/Georgis, and Sineshaw (1982), Ethiopian participants achieved an average accuracy of only 52% with a series of photographs developed by Ekman and Friesen (1976) that had American accuracy norms of 83%. Thus, it seems that although emotions are recognized at above-chance levels across cultures, there is also cultural variation in recognition accuracy.

The second goal of this article, therefore, is to examine the data to distinguish among the possible explanations described below for cultural variability in recognition rates. Several explanations have been forwarded regarding cultural universality and variation in recognition accuracy. Below we discuss some of the most prominent explanations that we attempt to examine in this meta-analysis.

### *Display Rules and Decoding Rules*

One explanation regarding cultural differences in emotion recognition suggests that cultural variation in the display and decoding rules serve as norms for the appropriateness of displaying and judging emotions (Ekman, 1972; Matsumoto, 1992a). These norms regulate the display of emotions in cases in which such display would disrupt social interaction (Ekman, 1972) as well as the decoding of emotions when recognizing these emotions might be potentially disruptive to social interaction (Matsumoto, 1992a). Thus, cultural variation in the accuracy of emotion recognition is attributed to norms that some cultures impose on members regarding the expression or recognition of certain emotions, particularly negative emotions. In support of this explanation, Matsumoto (1989) has demonstrated that some accuracy rates by country correlate with various dimensions of cultural variability.

### *Language*

Cultural variability in the accuracy of emotion recognition has also been attributed to differences in language. The words used to translate emotional concepts and labels may convey somewhat different meanings to cultures other than that from which the experiment originated (Harré, 1986; Mesquita et al., 1997). In a related vein, Matsumoto and Assar (1992) have suggested that the vocabulary of some languages might be better equipped to express emotional concepts than those of other languages.

Language attributes have also been implicated in explanations for cross-cultural variability in studies of emotion recognition from the voice rather than from still photographs of the face. In their review, Mesquita and Frijda (1992) noted that members of the same cultural group from which the vocal expressions were recorded often achieve the highest level of recognition accuracy. They suggested that general attributes of voice quality may be distracting for foreign listeners and may overshadow the actual vocal cues used to express emotions. Scherer, Banse, and Wallbott (2001) also argued that characteristic suprasegmental cues such as intonation contours and rhythm can vary across language families and may be difficult for outsiders to decode. Given the importance of language across these explanations, it is important to note the frequent overlap and thus confound between native language and cultural group membership.

### *In-Group Advantage*

Variability in cross-cultural emotion recognition has also been attributed to in-group familiarity. It is possible that recognition accuracy is higher when emotions are both expressed and perceived by members of the same cultural group. This tendency has been called *ethnic bias* by some theorists (e.g., Kilbride & Yarcower, 1983; Markham & Wang, 1996). The term *bias*, however,

suggests inaccuracy, and for that reason we prefer to use the term *in-group advantage*. Judging the emotions of other members of one's own social group more accurately is an increase in accuracy, rather than a bias. Thus, an in-group advantage in judging emotions might be associated with a match between the cultural and social background of the judge and target. The in-group advantage can be construed in terms of display and decoding rules because accuracy is higher when display rules and decoding rules are congruent. However, this advantage might not necessarily be due to social pressures or norms as suggested by Ekman (1972) and Matsumoto (1989). Rather, it is equally plausible that the in-group advantage might be associated with subtle stylistic differences across cultures in the method of expressing basic emotions.

Previous research examining the existence of an in-group advantage has yielded inconclusive results. For instance, in examining the recognition of facial expressions of emotion, Ekman (1972) did not find an in-group advantage for American and Japanese participants. Similarly, Boucher and Carlson (1980) and Kilbride and Yarczower (1983) found no evidence for an in-group advantage in their work with Americans both judging and being judged by Malaysians and Zambians, respectively. On the other hand, Albas, McCluskey, and Albas (1976) found that Canadians from European and Cree Indian backgrounds showed significantly better vocal emotion recognition of members of their own group. More recently, Markham and Wang (1996) found the same significant interaction effect with Chinese and Caucasian children. Using vocal expressions of emotion, Scherer et al. (2001) compared judgments of German actors' expressions of emotion by Germans as well as by members of eight other cultures. They found that Germans had the highest performance, and that emotion recognition accuracy was greater for foreign judges whose own language was closer to the Germanic language family. Similarly, in examining the accuracy of the judgment of positive and negative nonverbal behavior from different dynamic channels of communication, Rosenthal et al. (1979) also found evidence that the magnitude of in-group advantage was smaller for groups with greater cross-cultural exposure to the United States. In sum, previous evidence is mixed regarding an advantage in recognizing emotions expressed by members of the same cultural group as the perceiver.

### Nonverbal Channels of Emotional Expression

A third goal of this article is to explore whether emotion recognition within and across cultures varies depending on the particular nonverbal channel of the emotional expression. Much of the evidence considered in the debate about the universality of emotion recognition focuses on the recognition of still photographs of facial expressions (Ekman, 1994; Galati, Scherer, & Ricci-Bitti, 1997; Izard, 1994; Russell, 1994), but emotions are often conveyed via multiple channels of communication. Restricting investigations of emotion recognition to just the facial channel, albeit convenient and allowing for tighter experimental control, raises questions about ecological validity. As early as 1954, Bruner and Tagiuri called for studies of emotional judgments to move beyond the still facial photograph, saying,

Historically speaking, we may have been done a disservice by the pioneering efforts of those who, like Darwin, took the human face in

a state of arrested animation as an adequate stimulus situation for studying how well we recognize human emotion. (p. 638)

Fortunately, there has been excellent—although relatively sparse in comparison—experimental work examining emotion recognition from channels other than still facial photographs, including dynamic, expressive channels of communication. For example, the cross-cultural recognition of vocal expressions has been studied by Albas et al. (1976), Beier and Zautra (1972), Kramer (1964), Scherer et al. (2001), and others, and the cross-cultural recognition of emotional expressions through body movements has been studied by Gallois and Callan (1986), Rosenthal et al. (1979), Winkel-mayer, Exline, Gottheil, and Paredes (1978), and others. Moreover, recent work on the accuracy of judgments from nonverbal behavior suggests that even small amounts of dynamic nonverbal information can yield more accurate interpersonal judgments than judgments based on static information (Ambady, Bernieri, & Richeson, 2000; Ambady, Hallahan, & Conner, 1999; Wehrle, Kaiser, Schmidt, & Scherer, 2000). Consequently, it seems worthwhile to examine the accuracy of emotion recognition via multiple channels of communication rather than just the face.

### Differences in Expression Across Emotions

A fourth goal of this meta-analysis is to explore cultural universals and variability in emotion recognition in relation to the particular emotion being expressed. There has been active discussion in the emotion literature about which particular emotions are likely to be universally recognized. For instance, there has been some controversy (e.g., Izard & Haynes, 1988; Matsumoto, 1992b; Ricci Bitti, Brighetti, Garotti, & Boggi-Cavallo, 1989; Russell, 1991) about whether contempt should be included along with the original six basic emotions—anger, disgust, fear, happiness, sadness, and surprise—identified by Ekman (1972). Moreover, Russell (1994) has argued that broad dimensions such as valence and arousal are universal, whereas particular categories such as anger or surprise may vary by culture. This assertion has not gone unchallenged (e.g., Ekman, 1994; Izard, 1994). For this reason, it might be informative to examine cultural variability associated with the recognition of both the broad dimensions of positive and negative emotions as well as specific, discrete emotions.

### Moderators

Several interesting potential moderators of cross-cultural emotion recognition have been suggested in the literature (Mesquita & Frijda, 1992; Russell, 1994; Scherer, 1997; Scherer & Wallbott, 1994). We examine a number of these in this meta-analysis. They include (a) exposure between cultures, (b) majority or minority status in the culture, (c) attributes of studies, and (d) other demographic moderators.

#### *Exposure Between Cultures*

It is likely that the accuracy of emotion recognition will be moderated by the amount of exposure between the cultures of the judges and the targets. There has been considerable speculation that the style of emotional expression can differ across cultural groups such that individuals are better able to understand emotions that are expressed in their own familiar style (Albas et al., 1976;

Efron, 1941; Mehta, Ward, & Strongman, 1992; Scherer et al., 2001; Seaford, 1975). Rosenthal et al. (1979) conducted analyses suggesting that greater cross-cultural exposure is associated with a smaller degree of in-group advantage. Furthermore, experience and familiarity with faces from other cultures has been found to moderate the discrimination and recognition of same and different-race faces (O'Toole, Deffenbacher, Valentin, & Abdi, 1994; O'Toole, Peterson, & Deffenbacher, 1996). Thus, greater contact and familiarity should be associated with greater accuracy in recognizing emotions across cultures. We propose to test this hypothesis using a number of measures that operationalize the degree of exposure between cultural groups. First, we expect that cultural groups living together within a single nation versus across national borders have greater exposure to each other. Second, we examine cultural contact across nations using two available proxy measures: the physical proximity between groups and the degree of contact via telephone.

### *Majority or Minority Status in the Culture*

There has been a longstanding debate as to whether majority group members are more accurate at judging the emotions and behavior of minority group members and vice versa. Majority and minority groups may differ in their degree of exposure to each other—by sheer numerosity, many individuals experience greater exposure to members of the majority group. However, the “subordination hypothesis,” proposed by Henley (1977) further posited that intergroup power differences can lead minority group members to be more motivated to understand the nonverbal cues of majority group members than the reverse. Henley was particularly interested in gender relations, but this logic may apply to ethnic or regional minority groups as well. Evidence regarding the validity of the subordination hypothesis, however, is mixed (Deaux & LaFrance, 1998; J. A. Hall, 1998; J. A. Hall & Friedman, 1999; J. A. Hall, Halberstadt, & O'Brien, 1997; Hecht & LaFrance, 1998; Henley, 1977). For instance, J. A. Hall et al. (1997) conducted a meta-analysis of the relationship between nonverbal sensitivity and individual-level measures of subordination such as social class, gender attitudes, and leadership experience. They found little support for the subordination hypothesis because higher status was associated with greater rather than lesser sensitivity to nonverbal cues. Restricting our analysis to the judgment of emotions, we propose to compare the performance of majority and minority groups living together within the same nation.

### *Attributes of Studies*

There are many opinions about the optimal methodology for studying emotion recognition (e.g., Ekman, 1994; Izard, 1994; Matsumoto, 1992a; Rosenthal et al., 1979; Russell, 1994). We examine the impact of various attributes of research design and stimuli on the level of cross-cultural accuracy and the in-group advantage in emotion recognition.

### *Other Demographic Moderators*

For exploratory purposes, we analyze the effects of the year of study, participant age, and participant gender on the cross-cultural accuracy and in-group advantage in emotion recognition.

In sum, the overarching goal of this meta-analysis is to examine the evidence for the universality and cultural specificity of emotion recognition. In doing so, we examine emotion recognition from multiple channels of communication as well as the recognition of specific emotions.

## Method

### *Literature Search*

Several methods were used to locate the relevant studies:

1. First, an initial computer search of Psychological Abstracts (PsycLIT), Social Science Citation Index, Dissertation Abstracts International and the Educational Resources Information Center was conducted to retrieve documents containing the terms *culture*, *ethnicity*, *in-group*, *out-group*, or *nation*, and also *emotion* or *nonverbal*.

2. Articles referenced by usable articles from the first method were examined.

3. The Social Science Citation Index was used to check citations for any article that had been found through the other methods and was deemed usable for the study.

4. A manual search was conducted of several journals that had many usable articles found through the previous methods. Those journals included: *Cognition & Emotion* (1987–1999, Volumes 1–13), *Environmental Psychology and Nonverbal Behavior* (1976–1979, Volumes 1–3), *Journal of Personality and Social Psychology* (1987–1999, Volumes 53–77), *Journal of Nonverbal Behavior* (1979–1999, Volumes 4–23), *Journal of Cross-Cultural Psychology* (1970–1999, Volumes 1–30), *International Journal of Psychology* (1966–1999, Volumes 1–34), and *Motivation and Emotion* (1977–1999, Volumes 1–23).

5. Our own files were reviewed for preprints and unpublished manuscripts.

6. Finally, unpublished manuscripts were directly solicited. A letter was sent to authors of articles included in the study as of 1998 requesting any unpublished manuscripts relating to the topic of differences in emotional recognition across groups that differ in race or ethnicity. Forty percent of the authors responded to this request. Additionally, a general request was placed for such manuscripts via the Internet e-mail list for the Society for Personality and Social Psychology. Ten researchers responded to this general request.

Thus, every effort was made to obtain published as well as unpublished articles for this review.

To be included in this review, a study had to satisfy the following criteria:

1. A judgment of emotions must have been made in which there was an objective criterion for accuracy. The most typical criterion for accuracy was whether judges correctly identified the emotion that the poser had been asked to portray. We did not want to impose our own definition of the term *emotion* for the purpose of this review. Therefore, we allowed any emotion that individual researchers considered worthy to include in their experiment. We did, however, exclude judgments of gestures or personality traits. In addition to the standard anger, contempt, disgust, fear, happiness, positivity–negativity, sadness, and surprise, the emotions examined included acceptance, amusement, anticipation, anxiety, boredom, calmness, compassion, confusion, contentment, depression, doubt, embarrassment, excitement, flirtatiousness, friendliness, hostility, indifference, intensity, interest, irritation, love, neutral, pain, shame, sleepy, stubborn, submissive, and sympathy.

2. Emotions needed to be judged cross-culturally, that is, with both a judgment made by members of the same culture as the poser group (the in-group) and members of a cultural group other than that of the poser (the out-group). This was possible in two different ways. The same set of stimuli could be viewed both by members of the cultural group from which it originated and by members of a different cultural group. Alternatively, a

single group of participants could view stimuli both from members of their own cultural group and from members of a different cultural group. The same experimental methods and participant attributes such as age, gender, and so on needed to apply to both the in-group and out-group judgments.

3. Enough data must have been reported to determine the level of accuracy for the in-group and the out-group judgments. However, to be conservative with respect to the hypotheses regarding cultural differences, we included some studies even if the accuracy data were not reported but it was mentioned in the text that cultural differences in emotion recognition had been tested and no effect was found.<sup>1</sup>

4. Participants could not be members of clinical populations, such as individuals diagnosed with schizophrenia or mental retardation. However, "normal" control groups in clinical studies could be included if the data were reported separately.

Many studies were included that were not designed explicitly to examine cross-cultural comparisons. For example, many researchers have used Ekman and Friesen's (1976) "Pictures of Facial Affect" stimuli in their experiments in other countries. Similarly, others have used Scherer et al.'s (2001) German stimuli (e.g., Guidetti, 1991, in France). Such studies were included as long as participants were matched to the normative group in terms of experimental procedures, such as exposure time, and in terms of participant attributes, such as age and gender composition.

Note that we operationalized culture roughly to include different countries, subcultures, and ethnic groups, as well as regions within a country. An attempt was made to include multiple types of cultural and subcultural groups, although in practice sufficient data were only available to look at national, ethnic, and regional groups. Depending on the definition of *culture* one uses, gender might also be appropriate to include as a cultural group. However, excellent reviews of gender effects in nonverbal sensitivity have already been presented elsewhere (e.g., J. A. Hall, 1978, 1984).

### Coding Procedure

The following variables were coded for each study. Hillary Anger Elfenbein coded all variables, and either Nalini Ambady or a research assistant coded a subsample of at least 20% of the articles to calculate the interrater reliability. Coefficients listed are Spearman-Brown effective reliability values based on a single rater (Rosenthal & Rosnow, 1991).

*Attributes of the participants.* These included (a) the national, ethnic, or regional group of in-group members and that of out-group members. The in-group was the cultural group that judged emotional stimuli originating from its own group. The out-group was either the cultural group that judged emotional stimuli originating from another group or the cultural group of additional stimuli viewed by the in-group ( $r = 1.0$ ), (b) the number of participants overall (raters or decoders;  $r = 1.0$ ), (c) the percentage of female versus male participants ( $r = 1.0$ ), (d) the age of participants ( $r = 1.0$ ), (e) whether the socioeconomic status of the participants in the in-group and out-group was matched ( $r = 1.0$ ), and (f) whether the in-group was a majority or a minority group, for cultural groups living together within the same nation ( $r = 1.0$ ). The *majority group* was defined as the group in numerical majority in the society in which they live, although this also corresponded generally to higher status. When the in-group and out-group were not ethnic groups, but rather regional groups within a particular country, this particular variable was not applicable unless one of the regional groups was tested while residing in the other region (for example, Southern Italians living in Northern Italy). If both regional groups took part in the study while residing in their own respective regions, then both groups were considered majority groups and this variable was not included.

*Attributes of the study and stimuli.* These included (a) the channel of nonverbal communication used (e.g., still facial photos, voice, video without sound;  $r = .96$ ); (b) whether the study was published or unpublished ( $r = 1.0$ ); (c) the year of the study ( $r = 1.0$ ); (d) whether participants responded using multiple choice, a dimensional scale, or free response

( $r = 1.0$ ); (e) the particular researchers conducting the study ( $r = 1.0$ ); (f) whether the study had a balanced design—that is, whether every cultural group in the study judged emotions expressed by members of every other group ( $r = 1.0$ ); and (g) whether the recognizability of stimuli was validated prior to the study using a separate consensus sample of raters. The separate sample of raters could consist of participants from a previous study, providing that the researcher specified a criterion that poorly recognized stimuli were excluded from further use. For multiple sets of stimuli in a single study, this variable was coded, in particular, for the stimuli containing out-group members. This is because these out-group stimuli were used to calculate the cross-cultural judgment accuracy in such experiments (reliability was based on discussion between both authors until agreement was reached). Finally, studies included (h) whether the emotions in the stimuli were expressed spontaneously, posed, or imitated. Posing included instructions to pose a particular emotion or role-playing using emotional vignettes. Imitation included instructions to the poser to imitate a pose or photograph or to move particular muscles. Furthermore, imitation included stimuli selected using the criterion of consistency with particular poses, photographs, or muscular movements, even if the poser did not express the emotion at the time under such instructions (reliability was based on discussion between both authors until agreement was reached).

*Exposure between cultural groups.* These included (a) whether the in-group and out-group live together within a nation (within-nation comparisons) or across national borders (across-nation comparisons)—note that, for the purposes of this coding, Canada and the United States are considered together as a contiguous community ( $r = 1.0$ ); (b) the physical distance between the two countries, measured by the "as the crow flies" distance in kilometers between the capital cities of the two nations (where applicable; we found this information on an Internet Web site (<http://www.indo.com/distance/>) that makes calculations on the basis of the latitude and longitude positions of cities by using the method outlined by the United States Geological Survey); and (c) contact between cultures via telephone, measured by the total number of minutes of telephone communication between the in-group and out-group nations. If this number was not zero, it was scaled by dividing by the number of telephones in the cross-cultural decoder's nation. A square-root transformation normalized these values. We found this information in TeleGeography (1997). For items (b) and (c), in the case of multiple out-groups, the study codes were the average values for each of the country pairs.

### Computation of Effect Sizes

We computed three different effect size estimates. All effect sizes were computed by Hillary Anger Elfenbein and checked by Nalini Ambady. The Appendix contains two examples of the calculation of effect sizes.

*Interaction F test.* For fully balanced designs in which members of two or three different cultural groups judged stimuli from each group, we computed an *F* test for the Encoder  $\times$  Decoder interaction, as long as

<sup>1</sup> For the seven studies that otherwise met the criteria for inclusion but which did not report sufficient information to compare the accuracy of the groups, we determined how to include them on the basis of other information that was reported. For the four studies that reported null results for an in-group advantage *t* test or *F* test with one degree of freedom in the numerator (Ekman, 1972, Studies 1 and 2; Izard, 1971, Study 4; and McAndrew, 1986)—regardless of statistical power—we included these studies with an effect size of zero. For the three studies that reported either an omnibus *F* test or did not report about statistical tests that might have been conducted, we included these studies in the sign-test analysis only. These studies were assigned a negative value in the case of either null or negative results (Ekman et al., 1969, Study 2; Matsumoto, 1993; and Sorensen, 1975). Note that Matsumoto (1993) contained three in-group–out-group comparisons, so that three negative values were added to the sign test.

information on the variance or error term was included. Depending on the direction of the effect, this  $F$  test represents the extent to which in-group judgments were more accurate than out-group judgments, or vice versa. In the case of balanced experiments with three different groups, we calculated both the omnibus  $F$  test for the Encoder  $\times$  Decoder interaction, as well as the focused contrast  $F$  test using lambda weights of +2 for the three in-group judgments and -1 for the six out-group judgments.

*Percentage difference effect size.* Most of the studies reported data on emotion recognition for all cultural groups in terms of percentage accuracy. Therefore, percentage accuracy seemed to be a natural, inclusive, and unbiased index to use in the analysis of both cross-cultural emotion recognition and the in-group advantage.

For multiple-choice studies, it was necessary to correct these percentages for the degree of accuracy expected due to chance guessing. Because studies varied in the number of response choices, this correction was required so that the values from different studies could be directly comparable with each other using equivalent scales. In a multiple-choice response format, some component of accuracy can be attributed to chance guessing. It is important for this component to be consistent across studies, regardless of the number of response choices. The use of a small number of choices effectively limits the range of percentage accuracy. We used the standard correction formula,  $(\text{proportion correct} - (1/\text{number of choices})) / (1 - (1/\text{number of choices}))$ , (Nunnally & Bernstein, 1994). Conceptually, this means that we subtract the portion of the accuracy that is due to chance, and then we scale the remainder of the score by the new total score possible. Applying this guessing correction leads us to expect a corrected accuracy value of zero due to chance alone, and any number above zero is an indication of accuracy beyond chance guessing. For example, a value of 58% indicates an estimate that participants were correct approximately 58% of the occasions during which they did not merely guess at random among choices. It is preferable to use a formula that corrects not only for the degree of accuracy due to chance guessing, but also for the degree of accuracy due to response bias (Wagner, 1993). For example, a participant guessing *angry* for every choice would receive a score of 100% for anger stimuli without necessarily being able to discriminate anger correctly from other emotions. However, such a correction formula requires not only the conventional accuracy "hit rates" that researchers reported, but also confusion matrices listing the pattern of judgment errors (Wagner, 1993). Because such confusion matrices were provided in only a small portion of the studies,<sup>2</sup> we were unable to make this correction.

In the analysis of cross-cultural accuracy, the relevant effect size was the emotion recognition accuracy achieved by out-group members after the guessing correction. This value is an index of the extent to which emotions were accurately recognized cross-culturally.

In the analysis of cultural differences or the in-group advantage, the relevant effect size was the difference between the emotion recognition accuracy achieved by the in-group versus out-group members after the guessing correction. This value is an index of the in-group advantage, the extent to which emotions were recognized less accurately across cultural boundaries.<sup>3</sup> Whenever portions of the data necessary to calculate the cultural difference percentage were missing, we made the most conservative choice possible using the data available. For example, Ekman et al. (1969) reported the percentage of participants by cultural group who made the first or second most common choices. In one case, the correct response was not among the most common two responses made by out-group members, so we assumed that all of the remaining participants made the correct response. The purpose of such decisions was to ensure that the results of this meta-analysis did not overestimate the size of the in-group advantage. However, in these cases when conservative measures were taken, we excluded the sample from the universality analysis to be conservative with respect to that hypothesis as well. In the case of multiple dependent measures taken from the same group of participants, such as multiple-response types or multiple nonverbal channels, we separately

analyzed the results from each dependent measure and then pooled them to contribute a single value for each set of out-group participants.

*Effect size ( $r$ ).* We calculated effect sizes ( $r$ ) and significance levels for the in-group advantage from the information provided whenever possible, to represent the extent to which in-group judgments were more accurate than out-group judgments.

Calculating an effect size was only possible for about one third of all of the studies. To calculate the effect size when it was not already specified in the article, we needed (a) the mean accuracy of the two groups, (b) a variance or error term, and (c) the number of participants in each group. Most authors reported mean accuracy values and sample sizes for each group, but often there was no information about variance or error terms so that no tests of statistical significance or effect size could be calculated.

<sup>2</sup> Confusion matrices were reported for the following studies: Chan (1985); Gitter, Kozel, and Mostofsky (1972); Kirouac and Doré (1983, 1985); Kramer (1964); Mazurski and Bond (1993); Niit and Valsiner (1977); Ricci Bitti et al. (1989); Scherer et al. (2001); Van Bezooijen, Otto, and Heenan (1983); Wallbott (1991b); and Yik and Russell (1999).

<sup>3</sup> In order not to violate assumptions of independence, we were consistent in using the out-group judgment as the unit of analysis. This allowed us to make use of multiple comparisons that used the same set of in-group norms. Although there was only one in-group for a particular set of stimuli, multiple out-groups were relatively common. Therefore, we designated the in-group score as a fixed quantity and the out-group scores as random quantities. This meant that both the analysis of universals and cultural differences made use of a set of independent data points. In the case of the cultural differences, each independent data point represented the difference between a random quantity and a fixed reference. This was simple in most cases, but required care for studies in which multiple groups each judged stimuli from all other groups. In those cases, we made an in-group-out-group comparison separately for each set of decoders. These methods allowed us to make use of multiple comparisons per study without violation of the assumption of independence required for statistical inference. Although this choice is conservative with respect to statistical power—because the participant group rather than the individual participant is the unit of analysis—it is more sensible in this situation in which multiple experimental results are compared with norms established by a single set of participants.

Note that all analyses of effect size using this percentage accuracy are unweighted models, rather than adjusted for estimated variance as suggested by Hedges and Olkin (1985). Such weighting would be inappropriate for the percentage accuracy effect sizes. Because many studies do not provide measures of variance, the number of participants would otherwise provide a proxy for the variability of each result. This weighting gives more importance to larger studies under the assumption that such studies provide more reliable effect size estimates. However, within the current set of studies, there is large enough variance in the methods and populations that there is no reason to believe that the number of participants in particular provides a firm indication of the effect size's reliability. An unweighted analysis is more consistent with the decision that—in the absence of information on variability from individual studies—the primary unit of analysis is the out-group sample itself. Therefore, each out-group sample contributes equally as a single data point. The unweighted analysis prevented undue influence by two studies together accounting for nearly half the total participants—Rosenthal et al. (1979) with nearly 9,000 and Dickey and Knower (1941) with nearly 2,000. Note that we did weight the values by estimated variance when calculating heterogeneity statistics. The purpose of the heterogeneity statistic is to demonstrate whether it is plausible for a single normal distribution to have generated the observed data. Weighting the data is consistent with the spirit of answering this question, even if it may not be consistent with providing an estimate of the average effect size. The authors thank Don Rubin, chair of the statistics department at Harvard University, for these suggestions.

When an effect size was not already provided, we used one of two methods for its calculation. First, if a study reported an  $F$  test with one degree of freedom in the numerator, then the effect size could be calculated directly using the formula,  $r = \sqrt{F/(F + df)}$ , (Rosenthal & Rosnow, 1991). Alternatively, it could be calculated if the mean, standard deviation, and sample size were all reported for the in-group and out-group so that a two-sample  $t$  test could be calculated using the formulas,  $t = ((M1 - M2)/SD) * \sqrt{(n1*n2)/(n1 + n2)}$ , and,  $r = \sqrt{t^2/(t^2 + df)}$ , (Rosenthal & Rosnow, 1991). For many studies, the pooled standard deviation for the  $t$  test was calculated on the basis of the reported mean-squared error of an  $F$  test elsewhere in the analysis. Whenever possible, we estimated these error terms from the portion of the data that was based on the particular in-group and out-group members being included in the comparison. To maintain statistical independence of studies, only one effect size and significance test was included per group of participants. When multiple in-group-out-group comparisons could be calculated from a single study, the effect sizes were averaged into a single value, using an unweighted Fisher  $Z_r$  transformation. However, for the purposes of separately examining within-nation comparisons and across-nation comparisons, we calculated these effect sizes separately for studies that included both of these types of comparisons. In the case of multiple dependent measures taken from the same group of participants, such as multiple-response types or multiple nonverbal channels, the results from each dependent measure were separately analyzed and then pooled in order to contribute a single effect size value to the meta-analysis for each set of participants.

### Study Characteristics

A total of 87 articles, describing 97 separate studies, were identified that satisfied the criteria for inclusion described above. These articles included a total of 182 independent samples of cross-cultural out-groups for use in the present analysis. The out-group sample was the primary unit of analysis in this review. The 87 articles included a total of 22,148 participants or more—some articles did not list the number of participants—with a mean of 255 and a median of 100 participants per study. The studies represent a wide variety of cultural groups, including 42 different nations, 23 different ethnic groups, and 2 regional groups.

In the most common experimental design, used by 66 of the 97 different studies represented in the 87 articles, a single set of emotional stimuli was displayed for judgment to members of the culture from which it originated and to members of other cultures as well. A great benefit of this design is its convenience for researchers, as the creation and validation of stimuli is extremely time-consuming and this design requires only a single set of stimuli. However, such a design allows the possibility of confounding together various effects across cultures. For example, high cross-cultural accuracy could indicate highly recognizable expressions from the cultural group posing the stimuli, or good decoding skills from the cultural out-group judging the stimuli. Likewise, the in-group advantage can indicate good emotional communication skills in the cultural in-group or poor emotional communication skills in the cultural out-group.

Another similar design, used by 7 of the 97 different studies, was for members of a single culture to see multiple sets of emotional stimuli expressed by members of different cultures. This design controls for the level of decoding skill, as each person serves as his or her own control in the within-subjects design. This design is also relatively convenient for researchers, who need only recruit a single group of participants. However, the design also allows certain cultural effects to be confounded. In particular, cultures may differ in the extent to which their emotional expressions are easily recognized, and this can be confounded with the distinction between in-group and out-group cultures.

Twenty-one of the 97 different studies used a completely balanced  $n \times n$  design with multiple groups, in which members of each group viewed emotional stimuli from members of each other group. This type of study is

more difficult to conduct but allows for separate analysis of effects across recognizability of emotional expressions, skill in decoding emotional expressions, and the possible interaction between such “encoding” and “decoding” skills. Three of the 97 different studies were a hybrid between these types of designs. They contained multiple cultural groups as both judges and targets but did not use all of the  $n \times n$  conditions needed to be fully balanced.

Given the strengths and weaknesses of each design, it was important for the current meta-analysis to include studies from each type. Because the balanced design is the most robust with respect to the hypotheses regarding cultural differences, these studies offer the strongest source of evidence and are analyzed separately. However, because the other designs are more convenient and consequently more plentiful, they are also included in analyses exploring possible moderators.

Tables 1 and 2 contain information about the studies used in this analysis and the variables for which they were coded. Table 1 lists all studies for which percentage accuracy effects could be calculated for the analysis of cross-cultural accuracy and the in-group advantage. Table 2 lists all studies for which an effect size could be calculated for the analysis of the in-group advantage.

## Results

### Cross-Cultural Accuracy

Are emotional expressions accurately recognized across cultures at accuracy levels greater than that expected through chance guessing? To address this question, the following two different estimates were computed: (a) the sign of the study, that is, whether the cross-cultural judgments in each study achieved accuracy greater than chance and (b) the cross-cultural accuracy level, the average percentage accuracy of cross-cultural emotion judgments.

*Sign test.* In support of findings that emotions can be recognized across cultural boundaries at greater than chance levels, only 3.1% of the samples (5 out of 162 reporting sufficient information) showed cross-cultural accuracy that was no better than chance for any individual emotion.<sup>4</sup> Only one comparison (1 out of 162, or 0.6%) showed average cross-cultural accuracy less than or equal to that expected through chance guessing. This value would be 50% under the null hypothesis that cross-cultural emotion recognition consists of chance guessing alone. This corresponds to a binomial chance significance level of less than  $10^{-14}$  (one-tailed).

*Percentage accuracy effect size.* Analysis of the average cross-cultural percentage emotion recognition accuracy revealed  
(text continues on page 216)

<sup>4</sup> The five samples that included any evidence of cross-cultural emotional judgments at or below levels expected due to chance are as follows: (a) contempt and surprise in Gates (1923), in which ethnic minority Americans judged a White Anglo Saxon target; (b) contempt in Kramer (1964), in which Americans judged Japanese; (c) contempt, disgust, fear, happiness, sadness, and surprise in Sorensen (1975), in which the Bahinemo of New Guinea judged Americans; (d) contempt in Ricci Bitti et al. (1989), in which Southern Italians judged Americans; and (e) contempt in Russell, Suzuki, and Ishida (1993), in which Greeks and Japanese judged Americans. It is interesting to note that contempt is included in each of these five null results, as the universal recognition of contempt has recently been the subject of active discussion (e.g., Izard & Haynes, 1988; Matsumoto, 1992b; Russell, 1991; Ricci Bitti et al., 1989).

Table 1  
 Summary of Studies for Which Percentage Accuracy Effect Sizes Could Be Calculated ( $N = 168$ )

Study	Cultural group		<i>n</i>		Cross-cultural accuracy (%)	In-group advantage (%)	Border <sup>b</sup>	Consensus <sup>c</sup>
	In-group	Out-group	In-group <sup>a</sup>	Out-group				
Albas, McCluskey & Albas (1976)	Anglo Canadian	Canadian Cree	20	20	43.3	29.4	W	N
	Canadian Cree	Anglo Canadian	20	20	44.4	27.2	W	N
Bailey et al. (1998)	U.S.	Afro Caribbean	38	17	60.0	10.4	A	Y
	Caucasian American	African American	S	39	65.7	4.6	W	Y
Beier & Zautra (1972)	U.S.	Poland	52	55	43.6	8.4	A	Y
	U.S.	Japan	S	54	37.6	14.4	A	Y
Bhardwaj (1982)	Anglo North American	West Indian Canadian	P	30	72.6	5.3	W	Y
	Anglo North American	Quebecois Canadian	S	30	71.4	6.4	W	Y
	Anglo North American	Foreign Immigrant Canadian	S	30	69.1	8.8	W	Y
Biehl et al. (1997)	U.S., Japan	Sumatra	315	32	80.0	2.4	A	Y
	U.S., Japan	Vietnam	S	34	80.4	2.0	A	Y
	U.S., Japan	Poland	S	75	84.9	-3.2	A	Y
	U.S., Japan	Hungary	S	45	86.6	-5.3	A	Y
	Japanese American	Caucasian American	S	271	86.1	-1.4	W	Y
Boggi-Cavallo (1983)	U.S.	Italy	ND <sup>m</sup>	40	73.2	13.2	A	Y
Bormann-Kischkel et al. (1990)	North America	Germany	ND <sup>n</sup>	90	42.7	10.8	A	Y
Boucher & Carlson (1980)	U.S.	Malaysia	P	53	62.8	19.0	A	N
	Malaysia	U.S.	P	30	62.4	-10.1	A	Y
Buchman (1973)	Caucasian American	African American, Puerto Rican American	P	30	63.5	-3.8	W	N
	African American	Caucasian American, Puerto Rican American	P	30	59.8	2.5	W	N
	Puerto Rican American	Caucasian American, African American	P	30	59.8	-3.8	W	N
Caterina et al. (1999)	Italy	U.S. Germany	P	72	26.9	12.5	A	Y
Chan (1985)	U.S.	Hong Kong	ND <sup>o</sup>	124	65.8	15.3	A	Y
Collins (1996)	Caucasian American	African American	ND <sup>p</sup>	84	62.6	9.1	W	Y
Demertzis & Nowicki (1996)	Caucasian American	African American	75	27	68.3	7.8	W	Y
Dennis (1982)	Caucasian American	African American	20	20	50.5	6.0	W	Y
	Caucasian American	Hispanic American	S	20	54.3	2.3	W	Y
Dickey & Knower (1941)	U.S.	Mexico	1,244	616	56.6	-13.4	A	—
Ducci et al. (1982)	U.S.	Ethiopia	ND <sup>s</sup>	100	49.4	30.5	A	Y
Ekman (1972), Study 1	U.S.	Japan	80	80	—	0.0	A	N
	Japan	U.S.	80	80	—	0.0	A	N
Ekman (1972), Study 2	U.S.	Japan	—	—	—	0.0	A	N
	Japan	U.S.	—	—	—	0.0	A	N
Ekman (1972), Study 3	U.S.	New Guinea Fore	ND <sup>q</sup>	34	36.6	46.4	A	N
Ekman (1972), Study 4	U.S.	Japan	99	29	74.1	8.9	A	N
	U.S.	Brazil	S	40	80.0	3.0	A	N
	U.S.	Argentina	S	168	78.7	4.3	A	N
	U.S.	Chile	S	119	81.7	1.3	A	N
Ekman & Friesen (1971)	U.S.	New Guinea Fore	ND <sup>q</sup>	319	71.8	11.1	A	Y
Ekman et al. (1987)	U.S.	Estonia	30	85	80.4	6.0	A	—
	U.S.	Germany	S	67	76.9	9.5	A	—
	U.S.	Greece	S	61	79.0	7.4	A	—
	U.S.	Hong Kong	S	29	79.8	6.6	A	—
	U.S.	Italy	S	40	83.1	3.3	A	—
	U.S.	Japan	S	98	73.4	13.0	A	—
	U.S.	Scotland	S	42	84.6	1.8	A	—
	U.S.	Sumatra	S	36	70.4	15.9	A	—
	U.S.	Turkey	S	64	77.1	9.3	A	—
Ekman & Heider (1988)	Minangkabau Sumatra	U.S., Japan	P	57	80.8	9.9	A	Y
Ekman et al. (1972)	U.S.	New Guinea Dani	ND <sup>q</sup>	64	65.7	17.3	A	Y
Ekman et al. (1969), Study 1	U.S.	Borneo	ND <sup>q</sup>	15	42.0	41.0	A	Y
	U.S.	New Guinea Pidgin	ND <sup>q</sup>	18	44.6	38.4	A	Y
	U.S.	New Guinea Fore	ND <sup>q</sup>	14	43.8	39.2	A	Y
Evans et al. (1988)	U.S.	Australia	ND <sup>r</sup>	85	58.4	0.9	A	Y
Fernandez-Dols et al. (1993)	U.S.	Spain	ND <sup>s</sup>	19	91.0	9.0	A	Y



Stimuli <sup>d</sup>	Balanced design	Age <sup>e</sup>	Response <sup>f</sup>	Published <sup>g</sup>	% Female	Channel <sup>h</sup>	Research team <sup>i</sup>	Distance <sup>j</sup>	Telephone <sup>k</sup>	Majority judges minority <sup>l</sup>	SES match
PE	Y	2	M	Y	0	V	O			N	N
PE	Y	2	M	Y	0	V	O			Y	N
PE	N	2	M	Y	53	FP, V, PB	NW	2,317	23.5		
PE	N	2	M	Y	54	FP, V, PB	NW			N	N
PE	N	2	M	Y	—	V	O	7,179	1.0		
PE	N	2	M	Y	—	V	O	10,927	2.5		
I	N	2	M	N	50	FP	O			Y	N
I	N	2	M	N	50	FP	O			N	Y
I	N	2	M	Y	50	FP	O			N	N
I	N	2	M	Y	47	FP	EM	11,062	0.0		
I	N	2	M	Y	44	FP	EM	9,412	7.1		
I	N	2	M	Y	47	FP	EM	7,876	3.8		
I	N	2	M	Y	73	FP	EM	8,218	3.1		
I	N	2	M	Y	53	FP	EM			Y	Y
I	N	2	M	Y	50	FP	EM	7,244	4.2		
I	N	0.67	M	Y	—	FP	EM	6,141	5.0		
I	Y	2	M	Y	50	FP	EM	15,348	0.8		
I	Y	2	M	Y	0	FP	EM	15,348	5.6		
PE	Y	2	M	N	50	V, VD	O			Y	N
PE	Y	2	M	N	50	V, VD	O			N	N
PE	Y	2	M	N	50	V, VD	O			N	N
PE	N	2	M	Y	50	FP	EM	4,102	4.9		
I	N	2	M	Y	12	FP	IZ	13,175	13.6		
PE	N	0	M	N	—	FP, V, PB	NW			N	N
PE	N	2	M	N	—	FP	NW			N	Y
PE	N	2	M	N	100	V, VD	R			N	Y
PE	N	2	M	N	100	V, VD	R			N	Y
PE	N	1	M	Y	50	FP	O	3,029	19.4		
I	N	1	M	Y	50	FP	EM	11,536	0.0		
SP	N	2	M	N	—	VD	O	10,927	2.5		
SP	N	2	M	N	—	VD	O	10,927	4.1		
SP	N	2	M	N	—	VD	O	10,927	2.5		
SP	N	2	M	N	—	VD	O	10,927	4.1		
PE	N	2	M	Y	—	FP	EM	14,549	0.0		
I	N	2	M	Y	—	FP	EM	10,927	4.1		
I	N	2	M	Y	—	FP	EM	7,606	5.7		
I	N	2	M	Y	—	FP	EM	8,330	6.2		
I	N	2	M	Y	—	FP	EM	8,035	7.6		
I	N	1.5	M	Y	—	FP	EM	14,549	0.0		
I	N	2	M	Y	—	FP	EM	6,981	0.0		
I	N	2	M	Y	—	FP	EM	6,723	5.0		
I	N	2	M	Y	—	FP	EM	8,269	4.7		
I	N	2	M	Y	—	FP	EM	13,175	13.6		
I	N	2	M	Y	—	FP	EM	7,244	4.2		
I	N	2	M	Y	—	FP	EM	10,927	4.1		
I	N	2	M	Y	—	FP	EM	5,547	8.0		
I	N	2	M	Y	—	FP	EM	16,355	0.0		
I	N	2	M	Y	—	FP	EM	8,416	2.4		
I	N	2	M	Y	—	FP	EM	11,062	0.0		
I	N	1	M	N	16	FP	EM	14,549	0.0		
I	N	2	M	Y	—	FP	EM	14,549	0.0		
I	N	2	M	Y	—	FP	EM	14,549	0.0		
I	N	2	M	Y	—	FP	EM	14,549	0.0		
I	N	2	M	Y	—	FP	EM	14,549	0.0		
PE	N	2	M	Y	32	V, VD	R	15,945	6.8		
I	N	2	M	Y	—	FP	EM	6,101	3.6		

(table continues)

Table 1 (continued)

Study	Cultural group		<i>n</i>		Cross-cultural accuracy (%)	In-group advantage (%)	Border <sup>b</sup>	Consensus <sup>c</sup>
	In-group	Out-group	In-group <sup>a</sup>	Out-group				
Gabel & Wolwer (1992)	U.S.	Germany	ND <sup>s</sup>	15	72.5	19.7	A	Y
Gallois & Callan (1986)	Australian-born Australian	British-born and Italian-born Australian	P	82	36.0	1.3	W	N
Gates (1923)	Protestant U.S.	Irish, Jewish and Italian American	69	114	27.6	4.7	W	Y
Gitter, Black, & Mostofsky (1972a)	African American	Caucasian American	12	12	58.3	-5.8	W	N
Gitter, Black, & Mostofsky (1972b)	Caucasian American	African American	12	12	48.7	-1.4	W	N
Gitter, Kozel, & Mostofsky (1972)	African American	Caucasian American	40	40	57.9	-4.1	W	N
Gitter et al. (1971)	Caucasian American	African American	40	40	48.8	5.2	W	N
	Caucasian American	African American	P	183	61.2	-0.6	W	Y
	African American	Caucasian American	20	20	24.7	5.3	W	N
	American	African American	20	20	38.7	-13.3	W	N
Guidetti (1991)	Germany	France	ND <sup>t</sup>	50	59.2	8.1	A	Y
Habel et al. (2000)	U.S.	India	40	24	72.9	16.1	A	Y
	U.S.	Germany	S	29	89.7	-0.8	A	Y
Haidt & Keltner (1999)	U.S.	India	40	40	47.6	9.3	A	Y
Hatta & Nachshon (1988)	Japan	Israel	40	31	80.4	3.9	A	Y
Izard (1971), Study 1	U.S.	England	89	62	74.8	6.3	A	Y
	U.S.	Germany	S	158	77.8	3.2	A	Y
	U.S.	Sweden	S	41	81.0	0.0	A	Y
	U.S.	France	S	67	79.6	1.4	A	Y
	U.S.	Switzerland	S	36	76.7	4.3	A	Y
	U.S.	Greece	S	50	71.5	9.6	A	Y
	U.S.	Japan	S	60	60.4	20.6	A	Y
Izard (1971), Study 2	U.S.	Turkey	164	86	68.3	2.7	A	Y
	U.S.	India	S	68	51.7	19.3	A	Y
	U.S.	Japan	S	60	58.5	12.5	A	Y
Izard (1971), Study 4	Caucasian American	African American	85	66	—	0.0	W	Y
Kellogg & Eagleson (1931)	Caucasian American	African American	ND <sup>u</sup>	332	42.0	6.6	W	Y
Kilbride & Yarczower (1980)	U.S.	Zambia	16	6	46.0	45.0	A	Y
Kilbride & Yarczower (1983)	U.S.	Zambia	40	29	26.9	12.7	A	N
	Zambia	U.S.	35	102	26.9	-6.0	A	Y
Kirouac & Doré (1982)	Caucasian American	French Canadian	ND <sup>s</sup>	100	87.2	3.3	W	Y
Kirouac & Doré (1983)	Caucasian American	French Canadian	ND <sup>s</sup>	34	89.4	0.4	W	Y
Kirouac & Doré (1985)	Caucasian American	French Canadian	ND <sup>s</sup>	300	83.8	6.0	W	Y
Kramer (1964)	U.S.	Japan	P	27	47.0	10.0	A	Y
Kretsch (1968)	U.S.	Israel, Japan	P	85	18.1	4.7	A	N
	Israel	U.S., Japan	P	53	11.7	15.6	A	N
	Japan	U.S., Israel	P	52	12.9	5.9	A	N
Leung & Singh (1998)	U.S.	Hong Kong	ND <sup>v</sup>	60	72.9	16.3	A	Y
Mandal & Bhattacharya (1985)	U.S.	India	ND <sup>s</sup>	25	92.8	5.0	A	Y
Mandal & Palchoudhury (1989)	U.S.	India	ND <sup>s</sup>	100	93.3	3.7	A	Y
Mandal et al. (1986)	U.S.	India	ND <sup>s</sup>	100	77.4	14.9	A	Y
Markham & Wang (1996)	China	Australia	72	72	63.4	19.2	A	Y
Matsumoto & Assar (1992)	U.S., Japan	Indian	ND <sup>w</sup>	100	87.1	-6.6	A	Y
Matsumoto & Ekman (1988)	U.S. (non-Asian)	Japanese American, Japanese	P	111	80.3	-1.3	Both	N
Mazurski & Bond (1993)	U.S.	Australia	ND <sup>s</sup>	201	84.5	5.4	A	Y
McAndrew (1986)	U.S.	Malay and Chinese Malaysian	40	31	—	0.0	A	Y

Stimuli <sup>d</sup>	Balanced design	Age <sup>e</sup>	Response <sup>f</sup>	Published <sup>g</sup>	% Female	Channel <sup>h</sup>	Research team <sup>i</sup>	Distance <sup>j</sup>	Telephone <sup>k</sup>	Majority judges minority <sup>l</sup>	SES match
I	N	2	M	Y	40	FP	EM	6,723	5.0		
PE	N	2	M	Y	57	V, VD	O			Y	N
PE	N	0	FR	Y	—	FP	O			N	N
PE	Y	2	M	Y	50	FP	O			N	N
PE	Y	2	M	Y	50	FP	O			Y	N
PE	Y	2	M	Y	50	FP	O			N	N
PE	Y	2	M	Y	50	FP	O			Y	N
PE	N	2	M	Y	50	V, VD	O			Y	Y
PE	Y	0	M	Y	50	FP	O			N	Y
PE	Y	0	M	Y	50	FP	O			Y	Y
PE	N	2	M	Y	50	V	SH	877	4.7		
PE	N	2	M	Y	50	FP	O	12,038	5.7		
PE	N	2	M	Y	66	FP	O	6,723	5.0		
I	N	2	FR, M	Y	45	FP	O	12,038	5.7		
PE	N	2	M	Y	50	V	O	9,169	0.0		
I	N	2	M	Y	—	FP	IZ	5,913	8.0		
I	N	2	M	Y	—	FP	IZ	6,723	5.0		
I	N	2	M	Y	—	FP	IZ	6,646	5.4		
I	N	2	M	Y	—	FP	IZ	6,179	4.4		
I	N	2	M	Y	—	FP	IZ	6,667	7.7		
I	N	2	M	Y	—	FP	IZ	8,269	4.7		
I	N	2	M	Y	—	FP	IZ	10,927	4.1		
I	N	2	M	Y	—	FP	IZ	8,416	2.4		
I	N	2	M	Y	—	FP	IZ	12,038	5.7		
I	N	2	M	Y	—	FP	IZ	10,927	4.1		
I	N	0	M	Y	—	FP	IZ			N	Y
PE	N	0.5	FR	Y	49	FP	O			N	N
I	N	0	M	Y	50	FP	IZ	12,396	0.0		
I	Y	2	M	Y	80	FP	O	12,396	0.0		
I	Y	2	M	Y	29	FP	O	12,396	0.0		
I	N	2	M	Y	50	FP	EM			N	Y
I	N	2	M	Y	50	FP	EM			N	Y
I	N	2	M	Y	50	FP	EM			N	Y
PE	N	2	M	Y	0	V	O	10,927	2.5		
PE	Y	2	M	N	50	V	O	10,222	2.0		
PE	Y	2	M	N	50	V	O	9,343	7.9		
PE	Y	2	M	N	50	V	O	10,048	2.9		
I	N	1	M	Y	—	FP	EM	13,175	13.6		
I	N	2	M	Y	—	FP	EM	12,038	5.7		
I	N	2	M	Y	35	FP	EM	12,038	5.7		
I	N	2	M	Y	50	FP	EM	12,038	5.7		
PE	N	0	M	Y	50	FP	O	8,987	2.4		
I	N	2	M	Y	65	FP	EM	8,939	4.0		
I	N	2	M	N	50	FP	EM			Y	Y
I	N	2	M	Y	77	FP	EM	15,945	6.8		
I	N	2	M	Y	—	FP	EM	15,348	4.8		

(table continues)

Table 1 (continued)

Study	Cultural group		<i>n</i>		Cross-cultural accuracy (%)	In-group advantage (%)	Border <sup>b</sup>	Consensus <sup>c</sup>	
	In-group	Out-group	In-group <sup>a</sup>	Out-group					
McCluskey & Albas (1981)	Mexico	Canada	P	105	47.2	16.7	A	Y	
	Canada	Mexico	P	105	57.7	-16.0	A	Y	
McCluskey et al. (1975)	Mexico	Canada	P	70	32.7	10.5	A	N	
	Canada	Mexico	P	70	38.8	-12.1	A	N	
Mehta et al. (1992)	New Zealand Maori	Caucasian New Zealander	11	12	50.5	-6.9	W	N	
	Caucasian New Zealander	New Zealand Maori	15	16	23.0	15.1	W	N	
Niit & Valsiner (1977)	U.S.	Kirghizia	ND <sup>s</sup>	80	67.8	22.1	A	Y	
	U.S.	Estonia	ND <sup>s</sup>	70	77.1	12.7	A	Y	
Nowicki et al. (1998)	Caucasian American	African American	P	86	42.8	36.7	W	Y	
	African American	Caucasian American	P	43	81.1	-22.8	W	Y	
	Caucasian American	African American	P	54	47.2	30.0	W	Y	
	African American	Caucasian American	P	51	81.7	-20.6	W	Y	
Ricci Bitti et al. (1989)	Northern Italian	U.S., Southern Italian	P	40	31.0	9.2	A, W	Y	
	Southern Italian	U.S., Northern Italian	P	40	3.3	51.5	A, W	Y	
Ricci Bitti et al. (1979)	Northern Italian	Southern Italian	P	18	24.4	22.3	W	N	
	Southern Italian	Northern Italian	P	18	49.9	-30.5	W	N	
Ricci Bitti et al. (1980)	Northern Italian	Southern Italian	P	18	46.2	21.3	W	N	
	Southern Italian	Northern Italian	P	18	72.3	-25.3	W	N	
Rosenthal et al. (1979)	North America	Australia	4,544	645	53.4	6.6	A	Y	
	North America	Germany	S	25	53.6	5.5	A	Y	
	North America	Hong Kong	S	128	53.8	5.3	A	Y	
	North America	Ireland	S	71	54.0	5.3	A	Y	
	North America	Israel	S	123	50.8	8.3	A	Y	
	North America	Mexico	S	169	50.6	8.5	A	Y	
	North America	New Guinea	S	196	40.1	17.8	A	Y	
	North America	New Zealand	S	24	56.2	2.9	A	Y	
	North America	Singapore	S	69	49.2	7.2	A	Y	
	Caucasian American	Alaska Eskimo and Indian	427	61	46.9	11.1	W	Y	
Russell et al (1993)	North America	Greece	S	50	60.1	2.1	A	Y	
	North America	Japan	S	50	47.3	15.0	A	Y	
Scherer et al. (2001)	Germany	Switzerland	70	45	61.5	5.8	A	Y	
	Germany	England	S	40	60.5	6.8	A	Y	
	Germany	Netherlands	S	60	60.5	6.8	A	Y	
	Germany	U.S.	S	32	59.3	8.0	A	Y	
	Germany	Italy	S	43	59.3	8.0	A	Y	
	Germany	France	S	51	57.0	10.3	A	Y	
	Germany	Spain	S	49	52.0	15.3	A	Y	
	Germany	Indonesia	S	38	39.5	27.8	A	Y	
	Shimoda et al. (1978)	England	Italy, Japan	P	102	40.3	16.6	A	N
		Italy	England, Japan	P	32	34.9	23.1	A	N
Japan		England, Italy	P	30	50.6	-12.4	A	N	
Sogon & Masutani (1989)	Japan	U.S.	511	187	49.7	3.9	A	N	
Sorensen (1975), Study 1	U.S.	New Guinea Fore	ND <sup>d</sup>	68	28.0	52.7	A	Y	
	U.S.	New Guinea Bahinemo	ND <sup>d</sup>	71	0.0	76.6	A	Y	
Stokes (1984)	Caucasian American	African American	106	66	47.4	6.8	W	Y	
Streit et al. (1997)	U.S.	Germany	ND <sup>s</sup>	18	74.0	18.2	A	Y	
Sweeney et al. (1980)	U.S.	Japan	47	36	21.9	8.5	A	Y	
Toner & Gates (1985)	U.S.	Australia	ND <sup>s</sup>	64	85.6	4.3	A	Y	
Van Bezooijen et al. (1983)	Netherlands	Taiwan	48	40	30.2	31.8	A	N	
	Netherlands	Japan	S	41	26.1	35.9	A	N	
Wallbott (1991a)	U.S.	Germany	ND <sup>s</sup>	20	69.5	14.5	A	Y	
Wallbott (1991b)	U.S.	Germany	ND <sup>s</sup>	20	56.8	27.2	A	Y	
Wehrle et al. (2000)	U.S.	Germany	ND <sup>s</sup>	63	81.1	7.6	A	Y	
Winkelmayer et al. (1978)	U.S.	England	32	31	47.7	8.2	A	N	
	U.S.	Mexico	S	36	27.7	28.2	A	N	

Stimuli <sup>d</sup>	Balanced design	Age <sup>e</sup>	Response <sup>f</sup>	Published <sup>g</sup>	% Female	Channel <sup>h</sup>	Research team <sup>i</sup>	Distance <sup>j</sup>	Telephone <sup>k</sup>	Majority judges minority <sup>l</sup>	SES match
PE	Y	1	M	Y	0	V	O	3,599	2.0		
PE	Y	1	M	Y	0	V	O	3,599	1.4		
PE	Y	1	M	Y	0	V	O	3,599	2.0		
PE	Y	1	M	Y	0	V	O	3,599	1.4		
I	Y	2	M	Y	—	FP	O			N	N
I	Y	2	M	Y	—	FP	O			Y	N
I	N	1.5	M	Y	78	FP	EM	10,505	0.0		
I	N	1.5	M	Y	60	FP	EM	6,981	0.0		
PE	Y	2	M	Y	—	FP	NW			Y	Y
PE	Y	2	M	Y	—	FP	NW			N	Y
PE	Y	2	M	Y	—	FP	NW			Y	Y
PE	Y	2	M	Y	—	FP	NW			N	Y
PE	N	2	M	Y	50	FP	EM				
PE	Y	2	M	Y	50	FP	EM				
PE	Y	2	D	Y	50	VD	O			Y	Y
PE	Y	2	D	Y	50	VD	O			N	Y
PE	Y	2	M	Y	50	VD	O			Y	Y
PE	Y	2	M	Y	50	VD	O			N	Y
PE	N	1.5	M	Y	—	V, VD	R	15,945	6.8		
PE	N	1.5	M	Y	—	V, VD	R	6,723	5.0		
PE	N	1.5	M	Y	—	V, VD	R	13,175	13.6		
PE	N	1.5	M	Y	—	V, VD	R	5,453	11.1		
PE	N	1.5	M	Y	—	V, VD	R	10,927	1.4		
PE	N	1.5	M	Y	—	V, VD	R	3,029	19.4		
PE	N	1.5	M	Y	—	V, VD	R	14,549	0.0		
PE	N	1.5	M	Y	—	V, VD	R	14,075	5.9		
PE	N	1.5	M	Y	—	V, VD	R	15,544	11.3		
PE	N	1.5	M	Y	—	V, VD	R			N	N
I	N	2	M	Y	—	FP	EM	8,269	4.7		
I	N	2	M	Y	—	FP	EM	10,927	4.1		
PE	N	2	M	Y	67	V	SH	668	13.3		
PE	N	2	M	Y	50	V	SH	929	5.0		
PE	N	2	M	Y	63	V	SH	576	9.1		
PE	N	2	M	Y	44	V	SH	6,723	2.5		
PE	N	2	M	Y	56	V	SH	1,185	5.5		
PE	N	2	M	Y	51	V	SH	877	4.7		
PE	N	2	M	Y	43	V	SH	1,866	4.7		
PE	N	2	M	Y	71	V	SH	10,785	0.0		
PE	Y	2	M	Y	50	V, VD	O	5,515	2.7		
PE	Y	2	M	Y	50	V, VD	O	5,660	2.8		
PE	Y	2	M	Y	60	V, VD	O	9,731	0.7		
I	N	2	M	Y	35	VD	O	10,927	4.1		
I	N	—	M	Y	—	FP	EM	14,549	0.0		
I	N	—	M	Y	—	FP	EM	14,549	0.0		
PE	N	2	M	N	50	V, VD	R			N	Y
I	N	2	M	Y	33	FP	EM	6,723	5.0		
PE	N	2	M	Y	47	FP, PB	O	10,927	4.1		
I	N	2	M	Y	50	FP	EM	15,945	6.8		
PE	N	2	M	Y	50	V	O	9,479	0.0		
PE	N	1.5	M	Y	0	V	O	9,317	0.0		
I	N	2	M	Y	50	FP	EM	6,723	5.0		
I	N	2	M	Y	80	FP	EM	6,723	5.0		
I	N	2	M	Y	60	FP	EM	6,723	5.0		
SP	N	2	M	Y	0	VD	O	5,913	8.0		
SP	N	2	M	Y	0	VD	O	3,029	19.4		

(table continues)

Table 1 (continued)

Study	Cultural group		<i>n</i>		Cross-cultural accuracy (%)	In-group advantage (%)	Border <sup>b</sup>	Consensus <sup>c</sup>
	In-group	Out-group	In-group <sup>a</sup>	Out-group				
Wolfgang (1980), Study 1	Anglo Canadian	West Indian Canadian	P	—	59.2	18.1	W	Y
Wolfgang (1980), Study 2	West Indian Canadian	Anglo Canadian	W	—	58.0	-8.2	W	Y
	Anglo Canadian	West Indian Canadian	W	—	54.5	10.5	W	Y
Wolfgang & Cohen (1988)	Anglo Canadian	West Indian Canadian	P	79	65.0	10.5	W	Y
	Canadian	Israel	S	53	65.0	10.5	A	Y
	Canadian	Ethiopia	S	14	41.7	33.8	A	Y
	Anglo Canadian	Latin American Canadian immigrants	S	96	62.7	12.8	W	Y
Wolwer et al. (1996)	U.S.	Germany	ND <sup>s</sup>	21	74.2	18.0	A	Y
Yik et al. (1998)	China	Canada	50	50	42.6	-4.5	A	Y
	China	Japan	S	50	39.9	-1.8	A	Y
Yik & Russell (1999)	North America	Hong Kong	60	60	52.9	11.1	A	Y
	North America	Japan	S	60	56.3	7.7	A	Y

Note. Dashes indicate that information was not listed in the article. N = no; Y = yes.

<sup>a</sup> Number of participants in the comparison in-group; S = same participants as other out-groups in the same article; P = within-subjects design such that each participant judged in-group as well as out-group stimuli; ND = used normative data from another study, provides their own in-group comparison. Note that in-group accuracy is not the same for all studies using the same normative sample, as some studies used subsets of the stimulus materials. <sup>b</sup> Denotes whether studies use cultural groups that live across (A) national borders or within (W) the same nation. <sup>c</sup> Denotes whether studies used a consensus sample for the purpose of selecting emotional stimuli. <sup>d</sup> Method of eliciting emotional stimuli: PE = posed emotions; I = imitation of pose or theoretical construct; SP = spontaneous. <sup>e</sup> Participant age; 0 = child; 1 = teenager; 2 = adult or university student. <sup>f</sup> Type of response from participants: M = multiple choice; FR = free response; D = dimensional. <sup>g</sup> Published works included journal articles, books, book chapters, and manuscripts accepted for publication. Unpublished works included doctoral and master's-level theses, papers presented at conferences that were not included in the conference proceedings, and manuscripts not submitted or not yet accepted for publication. <sup>h</sup> Channel of nonverbal communication: V = voice;

that emotions can be judged across cultural boundaries at levels significantly greater than that expected due to chance alone. Such judgments were an average of 58.0% ( $SD = 19.6\%$ ) accurate,  $t(161) = 37.6$ ,  $p < 10^{-14}$ , one-tailed,  $r = .95$ , after correction for chance guessing. A stem-and-leaf diagram in Table 3 displays these effect sizes. Such a diagram is akin to a histogram turned on its side, the underlying data acting as part of the graphic to provide a visual display that preserves information (Rosenthal & Rosnow, 1991). Table 4 presents a statistical summary of the percentage accuracy effect sizes, including more information about measures of central tendency and variability.

### The In-Group Advantage

Is there evidence for an in-group advantage in emotion recognition whereby emotions are recognized more accurately when they are both expressed and perceived by members of the same national, ethnic, or regional group?

Four different analyses were computed: (a) the interaction  $F$  test, examining whether in-group judgments are more accurate than out-group judgments; (b) the sign of the study, that is, whether the in-group in each study achieved accuracy greater than the out-group; (c) the magnitude of the in-group advantage using the average percentage by which in-group judgments were more accurate than out-group judgments; and (d) the magnitude of the in-group advantage using the effect size ( $r$ ) expressing the extent to which in-group judgments were more accurate than out-group judgments.

**Interaction  $F$  test.** The interaction  $F$  test analysis supported the presence of an in-group advantage in emotion recognition. Table 5 lists the studies for which it was possible to perform an  $F$  test on

the in-group-out-group interaction term, which included 16 of the 21 studies that used a balanced design. The combined value of Stouffer's  $Z$ , a statistic that pools significance levels across studies (Rosenthal, 1991), for the predicted presence of an in-group advantage across the 16 studies was 2.17 ( $p < .02$ , one-tailed).

**Sign test.** The sign-test analysis supported the presence of an in-group advantage in emotion recognition; 80.2% of the results (146 out of 182) showed higher accuracy for judgments of in-group versus out-group members. This value would be 50% under the null hypothesis that emotion recognition rates are independent of the cultural match between encoder and decoder. This corresponds to a binomial chance significance level of less than  $10^{-14}$  (one-tailed), using the sample rather than the individual participant as the unit of analysis.

**Percentage accuracy effect size.** Analysis of the percentage accuracy effect size revealed an in-group advantage. Emotion recognition judgments made by participants from the same cultural group as the posers were an average of 9.3% more accurate than cross-cultural judgments ( $SD = 14.2\%$ ),  $t(167) = 8.5$ ,  $p < 10^{-14}$ , one-tailed,  $r = .55$ . This analysis uses the 168 samples for which the in-group advantage could be calculated, with the sample rather than individual participant as the unit of analysis. These percentage effect sizes are displayed in a stem-and-leaf diagram in Table 6. A statistical summary of the percentage accuracy effect sizes showing an in-group advantage, including more information about measures of central tendency and variability (Rosenthal, 1995) is presented in Table 4. In addition, Table 4 lists analyses of the heterogeneity of the in-group advantage across studies, using Rosenthal and Rubin's (Rosenthal, 1991; Rosenthal & Rubin, 1982) procedures for examining effect sizes expressed as differ-

Stimuli <sup>d</sup>	Balanced design	Age <sup>e</sup>	Response <sup>f</sup>	Published <sup>g</sup>	% Female	Channel <sup>h</sup>	Research team <sup>i</sup>	Distance <sup>j</sup>	Telephone <sup>k</sup>	Majority judges minority <sup>l</sup>	SES match
I	N	2	M	N	—	FP	O			Y	N
I	Y	2	M	N	—	FP	O			N	N
I	Y	2	M	N	—	FP	O			Y	N
I	N	2	M	Y	73	FP	O			Y	N
I	N	2	M	Y	100	FP	O	8,991	0.0		
I	N	2	M	Y	0	FP	O	11,140	0.0		
I	N	2	M	Y	51	FP	O			N	N
I	N	2	M	Y	29	FP	EM	6,723	5.0		
SP	N	2	D, FR	Y	50	FP	O	10,476	0.8		
SP	N	2	D, FR	Y	50	FP	O	2,103	2.1		
I	N	1.75	M	Y	50	FP	EM	13,175	13.6		
I	N	1.75	M	Y	50	FP	EM	10,927	4.1		

Note (continued). FP = facial photograph; PB = photographs of the body; VD = video. <sup>i</sup> Research teams, including borrowed stimulus materials: EM = Ekman and/or Matsumoto; IZ = Izard; NW = Nowicki; R = Rosenthal; SH = Scherer; O = Other. <sup>j</sup> Distance is the number of kilometers “as the crow flies” between capital cities. <sup>k</sup> Telephone communication is number of minutes per year telephone traffic between the in-group and out-group nations, if not zero then scaled by dividing by the number of telephones in the decoder’s nation (square-root transformed). <sup>l</sup> Denotes whether within-nations sample consisted of members of the majority group judging emotional stimuli expressed by minority group members (Y) versus members of a minority group judging emotional stimuli expressed by members of the majority group (N). <sup>m</sup> Norm based on Ekman et al. (1987). <sup>n</sup> Norm based on Bullock & Russell (1984). <sup>o</sup> Norm based on Izard (1971). <sup>p</sup> Normed based on Nowicki (1997). <sup>q</sup> Norm based on Ekman (1972), Study 4. <sup>r</sup> Norm based on Rosenthal et al. (1979). <sup>s</sup> Norm based on Ekman & Friesen (1976). <sup>t</sup> Norm based on Scherer et al. (2001). <sup>u</sup> Norm based on Gates (1923). <sup>v</sup> Norm based on McAlpine et al. (1991). <sup>w</sup> Norm based on Matsumoto & Ekman (1988).

ences in proportions. These formulas were adapted from Hedges’s (Hedges, 1982; Hedges & Olkin, 1985) formulas for effect sizes expressed in terms of Hedges’s *d*.

Because the percentage accuracy effect size made use of studies with unbalanced designs—that is, studies in which every group did not judge emotions expressed by members of every other group—we conducted an additional analysis. The purpose of the second analysis was to determine whether these percentage accuracy effect size results would be consistent with the interaction *F* test results obtained above with balanced design experiments. The goal of this additional analysis was to combine a subset of the unbalanced studies, using similar participant groups, into a single analysis to simulate a balanced design. If this analysis was consistent with the results using balanced studies, we could then feel comfortable making use of the larger group of unbalanced studies to test the in-group advantage hypothesis.

For this purpose, we combined all unbalanced studies involving one Caucasian group (North America, Europe, Australia, New Zealand) and one East Asian group (China, Hong Kong, Taiwan, Japan). The subset of samples is summarized in Table 7. The rationale for combining studies in this way is to control for any cultural differences in overall emotional expression and perception ability between Western and Asian cultures, as previously theorized and explored (e.g., Matsumoto, 1989, 1992a). Using this subset, we first replicated the above results that used the percentage accuracy effect size. In the subset, emotion recognition judgments made by participants from the same cultural group as the posers were an average of 13.4% (*SD* = 9.5%) more accurate than cross-cultural judgments,  $t(17) = 6.0, p < 10^{-5}$ , one-tailed,  $r = .82$ , which is consistent with the analysis of the larger sample of studies. Furthermore, we could combine these samples into a 2

(encoder: Caucasian and East Asian) × 2 (decoder: Caucasian and East Asian) analysis of variance (ANOVA). This combination treats each sample as if it were an individual participant in a single study with  $N = 18$ . The *F* test on the ANOVA interaction term represented the in-group advantage hypothesis, using an unweighted means analysis that reflects differences in the number of studies across conditions (Rosenthal & Rosnow, 1991). After accounting for the main effects of encoder skill, decoder skill, and clarity level, we obtained a significant predicted interaction term,  $F(1, 16) = 5.1, p < .02$ , one-tailed,  $r = .49$ , which indicates that these unbalanced samples give results consistent with those obtained earlier using fully balanced designs.

*Effect size (r).* Whenever possible, we calculated an effect size (*r*) as an index of the extent to which in-group judgments were more accurate than out-group judgments on the same emotion recognition task. Such effect sizes could be calculated for only about one third of the usable studies. These studies are likely to be somewhat unrepresentative, given that they exclude almost all of the classic work of Ekman and colleagues (e.g., Ekman, 1972; Ekman et al., 1969) and those who sought to replicate or extend this work using the same stimuli (e.g., Kirouac & Doré, 1983).

Keeping this limitation in mind, the effect-size analysis also indicated the presence of an in-group advantage. Members of the in-group culture appeared to achieve more accurate emotion recognition than out-group members. The unweighted average effect size was .25 (95% confidence interval (CI) .15 to .34) for the in-group advantage. Thus, this analysis also indicated the presence of an in-group advantage in emotion recognition, even with the smaller sample of studies. It was possible to calculate an individual significance value for each study that allowed us to calculate an effect size. To be conservative with respect to the hypothesis of

Table 2

Summary of Studies, Groups, and Nonverbal Channels for Which an Effect Size ( $r$ ) Could Be Calculated for the In-Group Advantage ( $N = 48$ )

Study	In-group	Out-group	Border	Channel	$Z_r^a$	$p$
Albas et al. (1976)	Caucasian North American and Canadian Cree	Caucasian North American and Canadian Cree	Within	V	.861	.000
Bailey et al. (1998) <sup>b</sup>	U.S.	Afro Caribbean	Across	V, FP, PB	.276	.023
Bhardwaj (1982)	Caucasian American Anglo Canadian	African American West Indian, Quebecois, and foreign immigrant Canadian	Within	V, FP, PB FP	.141 .273	.113 .006
Bond et al. (1990)	U.S. and Jordan	U.S. and Jordan	Across	FV	.417	.000
Bormann-Kischkel et al. (1990)	Canada	Germany	Across	FP	.537	.000
Buchman (1973)	Caucasian American, African American, Puerto Rican American	Caucasian American, African American, Puerto Rican American	Within	SV, V	-.069	-.260
Collins (1996)	Caucasian American	African American	Within	V, FP, PB	.310	.000
Demertzis & Nowicki (1996)	Caucasian American	African American	Within	FP	.277	.003
Dennis (1982)	Caucasian American, African American	Caucasian American, African American	Within	SV, FV	.219	.049
Dickey & Knower (1941)	U.S.	Mexico	Across	FP	-.748	.000
Ekman (1972), Study 1	U.S. and Japan	U.S. and Japan	Across	SV	.000	.500
Ekman (1972), Study 2	U.S. and Japan	U.S. and Japan	Across	SV	.000	.500
Gallois & Callan (1986)	Australian-born Australians	British-born and Italian-born Australians	Within	SV, FV	.227	.022
Gates (1923)	U.S. Protestant	Irish, Jewish and Italian American	Within	FP	.511	.000
Gitter, Black, & Mostofsky (1972a)	Caucasian American, African American	Caucasian American, African American	Within	FP	-.155	-.149
Gitter, Black, & Mostofsky (1972b)	Caucasian American, African American	Caucasian American, African American	Within	FP	.026	.371
Gitter, Kozel, & Mostofsky (1972)	Caucasian American	African American	Within	FP, V, SV, FV	-.108	-.074
Gitter et al. (1971)	Caucasian American, African American	Caucasian American, African American	Within	FP	-.062	-.139
Habel et al. (2000)	U.S.	India	Across	FP	.232	.014
C. W. Hall et al. (1996)	U.S.	China	Across	FP, SV	.071	.076
Hatta & Nachshon (1988)	Japan	Israel	Across	V	.096	.213
Izard (1971), Study 3	U.S.	France	Across	FP	.339	.000
Izard (1971), Study 4	Caucasian American	African American	Within	FP	.000	.500
Kilbride & Yarczower (1980)	U.S.	Zambia	Across	FP	.996	.000
Kilbride & Yarczower (1983)	U.S., Zambia	U.S., Zambia	Across	FP	.059	.199
Kretsch (1968)	U.S., Israel, Japan	U.S., Israel, Japan	Across	V	.366	.000
Leung & Singh (1998)	U.S.	Hong Kong	Across	FP	.518	.000
Machida (1986)	Caucasian American and Mexican American	Caucasian American and Mexican American	Within	SV	.078	.335
Markham & Wang (1996)	China	Australia	Across	FP	.658	.000
McAndrew (1986)	U.S.	Malay and Chinese Malaysians	Across	FP	.000	.500
McCluskey & Albas (1981)	Mexico, Canada	Mexico, Canada	Across	V	.021	.381
McCluskey et al. (1975)	Mexico, Canada	Mexico, Canada	Across	V	-.043	-.307
Mehta et al. (1992)	New Zealand Maori, Caucasian New Zealanders	New Zealand Maori, Caucasian New Zealanders	Within	FP	.138	.161
Nowicki et al. (1998)	Caucasian American, African American	Caucasian American, African American	Within	FP	.154	.010
Ricci Bitti et al. (1989)	Italy Northern and Southern Italians	U.S. Northern and Southern Italians	Across Within	FP FP	1.364 .388	.000 .000



Table 2 (continued)

Study	In-group	Out-group	Border	Channel	Z <sub>r</sub> <sup>a</sup>	p
Rosenthal et al. (1979) <sup>c</sup>	North America	Australia, Germany, Hong Kong, Ireland, Israel, Mexico, New Guinea, New Zealand, Singapore	Across	FV, SV, V	.344	.000
	Caucasian American	Non-White U.S., Alaska Eskimo and Native American	Within	FV, SV, V	.218	.000
Russell et al. (1993)	North America	Greece and Japan	Across	FP	.133	.061
Scherer et al. (2001)	Germany	Switzerland, England, Netherlands, U.S., Italy, France, Spain, Indonesia	Across	V	.405	.000
Shimoda et al. (1978)	England, Italy, Japan	England, Italy, Japan	Across	FV	.381	.000
Sprouse et al. (1995)	Caucasian American	U.S. non-Caucasian	Within	V, FP, PB	-.134	-.083
Stokes (1984)	Caucasian American	African American	Within	FV, SV, V	.286	.000
Strong (1978)	Caucasian American	African American	Within	FP	.213	.014
Sweeney et al. (1980)	U.S.	Japan	Across	FP	.686	.000
Winkelmayer et al. (1978)	U.S.	England, Mexico	Across	SV	.369	.000
Wolfgang & Cohen (1988)	Anglo Canadian	West Indian Canadian, Immigrants to Canada from Latin America	Within	FP	.519	.000
	Canada	Israel, Ethiopia	Across	FP	.548	.000
Yilk & Russell (1999)	North America	North America, China, Japan	Across	FP	.197	.005

Note. V = voice; FP = facial photograph; PB = photographs of body; SV = silent video; FV = full video with sound. <sup>a</sup>Z<sub>r</sub> is the effect size (r) after a Fisher transformation for normality. <sup>b</sup>In order to preserve statistical independence, the separate effect sizes for across-nation and within-nation samples were pooled together in the overall analysis. <sup>c</sup>Following the analysis conducted in Rosenthal et al. (1979), we obtained an effect size by comparing the foreign versus domestic status of individual samples' scores and their rank order. Our result differs from that presented in Rosenthal et al. (1979) only in that we have excluded all American occupational or age samples, such as children, for which there is no corresponding foreign sample.

cultural differences, we truncated all individual significance levels so that no single study was more significant than .05, as suggested by Rosenthal (1991). In spite of this truncation and the limitation of being able to include only about one third of the studies, the pooled test of overall significance from these individual studies indicated strong evidence for an in-group advantage. The com-

bined Stouffer's Z for this predicted effect was 7.06 ( $p < 10^{-11}$ , one-tailed). The effect sizes are shown in a stem-and-leaf display in Table 8. A statistical summary of these effect sizes, including more information about measures of central tendency, variability, and heterogeneity is presented in Table 4.

Channel of Communication

Did the cross-cultural accuracy of emotion recognition vary according to the nonverbal channel of expression? The studies in this meta-analysis included the channels of the face, the body, the voice, silent video, and combinations of these channels. We performed analyses examining differences across these channels, using the sample rather than the individual participant as the unit of analysis. Because many researchers made use of multiple channels in the same studies, we used a least squares multiple regression, with dummy codes representing whether specific nonverbal channels were present in the study, rather than a categorical model. This allowed us to consider studies with multiple channels even when the data were not reported separately by channel. For example, a study combining results using the voice as well as video would receive codes of +1 for voice, +1 for video, 0 for photographs of the body, and 0 for photographs of the face.

This analysis revealed some differences in cross-cultural emotion recognition accuracy across nonverbal channels. In particular, for the sample of studies with cultural groups spanning national borders, cross-cultural accuracy was lower for studies that used tone of voice ( $n = 36$ ) than it was for other ( $n = 81$ ) channels ( $b = -14.4\%$ , 95% CI =  $-23.2\%$  to  $-5.6\%$ ,  $p < .005$ ). Overall,

Table 3  
Stem and Leaf Plot of Mean Proportion Cross-Cultural Accuracy of Emotion Recognition (N = 162)

Stem	Leaf
.9	1, 2, 3
.8	5, 6, 6, 7, 7, 9, 9
.8	0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 3, 3, 4, 4, 4,
.7	6, 6, 7, 7, 7, 7, 8, 9, 9, 9, 9
.7	0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4
.6	5, 5, 5, 5, 5, 7, 8, 8, 9, 9
.6	0, 0, 0, 0, 1, 1, 2, 2, 2, 2, 3, 3
.5	6, 6, 6, 6, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9
.5	0, 0, 0, 0, 0, 1, 2, 2, 3, 3, 3, 4, 4, 4
.4	6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 9
.4	0, 0, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4
.3	5, 6, 7, 8, 8, 9, 9
.3	0, 1, 2, 4,
.2	6, 6, 6, 6, 7, 7, 8
.2	1, 3, 4, 4
.1	8
.1	1, 2
.0	
.0	0, 3

Table 4  
*Statistical Summary of Effect Sizes in Cross-Cultural Accuracy and In-Group Advantage in Emotion Recognition*

Statistic	Cross-cultural accuracy	In-group advantage		
		Overall	Across nations	Within nations
<b>Overall</b>				
No. of articles	80	87	60	31
Participants ( <i>N</i> )	20,448	22,418	17,619	4,918
Total no. of comparisons across groups	162	182	128	56
Proportion supporting hypothesis (%)	99.4	80.2	87.5	64.3
<b>Significance tests<sup>a</sup></b>				
<i>p</i> for binomial sign test	<.00001	<.00001	<.00001	<.02
<i>p</i> for <i>t</i> test of percentage difference	<.00001	<.00001	<.00001	<.02
Stouffer's <i>Z</i>		7.06	5.91	3.74
<i>p</i> for Stouffer's <i>Z</i>		<.00001	<.00001	<.0001
<b>Central tendency (percentage effect size)</b>				
<i>M</i> (%) effect size	58.0	9.3	11.6	4.6
No. of studies included in calculation	162	168	122	48
<i>t</i> based on percent differences	37.64	8.52	9.00	2.11
Effect size ( <i>r</i> ) based on <i>t</i> statistic	.95	.55	.63	.29
Proportion > 0.00 (%)	99.4	80.3	85.2	64.6
95% confidence interval	54.9–61.0	7.2–11.5	9.0–14.1	0.2–8.9
<b>Variability (percentage difference)</b>				
Maximum	93.3%	76.6%	76.6%	45.2%
Quartile 3	74.0	15.1	16.1	10.5
Median (%)	58.8	7.6	8.5	5.0
Quartile 1	45.7	2.2	3.6	–3.2
Minimum (%)	0.0	–30.5	–16.0	–30.5
<i>SD</i> (%)	19.6	14.2	14.2	15.0
Heterogeneity test		245.6	213.6	43.9
<i>p</i> for heterogeneity test		<.00001	<.00001	.48
<b>Central tendency (effect size <i>r</i>)</b>				
Mean effect size ( <i>r</i> )		.25	.30	.18
No. of studies included in calculation		48	27	22
Proportion > 0.00 (%)		77.1	81.5	72.7
95% confidence interval		.15–.34	.15–.43	.07–.28
<b>Variability (effect size <i>r</i>)</b>				
Maximum		.88	.88	.70
Quartile 3		.38	.48	.27
Median		.22	.33	.18
Quartile 1		.02	.06	–.02
Minimum		–.63	–.63	–.15
<i>SD</i>		.32	.37	.24
Heterogeneity test		719.9	651.3	77.1
<i>p</i> for heterogeneity test		<.00001	<.00001	<.00001

*Note.* Across-nation and within-nation comparisons do not always add to the total because some studies contained both within-nation and across-nation comparisons.

<sup>a</sup>Significance levels are listed as one-tailed in the direction supporting the hypothesized presence of cross-cultural accuracy as well as an in-group advantage in emotion recognition.

dynamic channels (tone of voice and video,  $n = 39$ ) were less accurately recognized across cultures than the static channels (photographs of face or body,  $n = 78$ ) in these studies ( $b = -6.6\%$ , 95% CI =  $-9.8\%$  to  $-3.4\%$ ,  $p < 10^{-4}$ ). This is consistent with arguments by theorists that vocal stimuli are more complex and less stylized than facial stimuli (Galati, et al. 1997). For the sample of studies using cultural groups contained within a national border, results did not differ by nonverbal channel,  $F(3, 43) = 0.4$ ,  $p > .7$ , or by dynamic versus static status of the channel,  $t(46) = -0.2$ ,  $p > .8$ . In terms of cultural differences, the size of the in-group advantage did not differ significantly across specific channels,  $F(3, 118) = 1.1$ ,  $p > .3$ , for cultural groups across national borders;

$F(3, 44) = 1.6$ ,  $p > .2$ , for cultural groups within a nation. Dynamic channels ( $n = 43$ ) showed a marginally smaller magnitude of in-group advantage than the static channels ( $n = 79$ ) for cultural groups across national borders,  $t(120) = -1.7$ ,  $p < .10$ , although this is not significant for cultural groups within national borders,  $t(46) = -0.2$ ,  $p > .8$ .

*Nonverbal channel* is an element of choice that particular research teams make in their experimental design, and consequently the nonverbal channel and research team can be linked. For example, Ekman and his colleagues (e.g., Ekman, 1972; Matsumoto & Ekman, 1988) have worked almost exclusively with still facial photographs, whereas Scherer and his colleagues (e.g., Scherer et

Table 5  
 Summary of *F* Tests on Encoder × Decoder Interaction in Studies With Balanced Designs (*N* = 16)

Study	Groups	Interaction <i>F</i> test	Direction of in-group advantage	<i>r</i> <sub>contrast</sub>	<i>p</i>
Albas et al. (1976)	Anglo Canadian and Canadian Cree	$F(1, 76) = 71.65$	yes	.70	<.00001
Bond et al. (1990)	Jordan, U.S.	$F(1, 188) = 34.92$	yes	.40	<.00001
Buchman (1973)	African American, Caucasian American, Puerto Rican American	omnibus $F(4, 261) = 1.02$ contrast $F(1, 261) = 2.08$	no	-.09	-.151
Gitter et al. (1972a)	African American, Caucasian American	$F(1, 40) = 1.18$	no	-.17	-.285
Gitter et al. (1972b)	African American, Caucasian American	$F(1, 144) = 0.11$	yes	.03	.739
Gitter et al. (1971)	African American, Caucasian American	$F(1, 72) = 0.15$	no	-.05	-.703
Kilbride & Yarczower (1983)	U.S., Zambia	$F(1, 202) = 0.83$	yes	.06	.364
Kretsch (1968)	Israel, Japan, U.S.	omnibus $F(4, 561) = 30.06$ contrast $F(1, 561) = 100.87$	yes	.39	<.00001
Machida (1986)	Caucasian American, Mexican American	$F(1, 28) = 0.18$	yes	.08	.671
McCluskey & Albas (1981)	Canada, Mexico	$F(1, 196) = 0.03$	yes	.01	.868
McCluskey et al. (1975)	Canada, Mexico	$F(1, 276) = 0.48$	no	-.04	-.490
Mehta et al. (1992)	Caucasian New Zealanders, New Zealand Maori	$F(1, 50) = 0.57$	yes	.11	.454
Nowicki et al. (1998), Study 1	African American, Caucasian American	$F(1, 125) = 5.98$	yes	.21	.016
Nowicki et al. (1998), Study 2	African American, Caucasian American	$F(1, 98) = 6.03$	yes	.24	.016
Ricci Bitti et al. (1989)	Northern Italy, Southern Italy	$F(1, 156) = 7.85$	yes	.22	.006
Shimoda et al. (1978)	England, Italy, Japan	omnibus $F(4, 483) = 21.17$ contrast $F(1, 483) = 81.07$	yes	.38	<.00001

Note. All significance levels are listed as one-tailed in the direction supporting the hypothesized presence of an in-group advantage in emotion recognition.

al., 2001) have worked exclusively with the voice. Therefore, it seemed worthwhile to examine the effect of nonverbal channel within studies, in addition to examining it across studies, to reduce any possible confound between the experimental research group and the nonverbal channel.

One particular set of studies included in this meta-analysis—those using the Profile of Nonverbal Sensitivity test (PONS), developed by Rosenthal et al. (1979)—allowed us to examine multiple results reported separately by nonverbal channel within the same study. These channels included the face, the body, the figure (face and body), content-filtered speech, random-spliced speech, and combinations of these channels. Detailed results are reported in Rosenthal et al. (1979) for the total of the 47 separate cross-cultural samples taking the PONS test, which is the unit of analysis. Note that these 47 samples are pooled by country to provide 10 separate out-groups in the other analyses in the present article. After correcting these accuracy values for chance guessing, we analyzed both the cross-cultural accuracy and the in-group advantage using a repeated-measures ANOVA. For each sample, we used the in-group accuracy based on the most appropriate American normative group, by age and professional status, to calculate the level of in-group advantage.

Looking first at cross-cultural accuracy, the omnibus *F* test for nonverbal channel was highly significant,  $F(10, 460) = 474.5$ ,  $p < 10^{-14}$ , suggesting strongly that the ability to recognize cross-culturally the broad dimension of positivity and negativity varies according to the nonverbal channel. Cross-cultural accuracy at above-chance levels varied from 19.6% for content-filtered sound to 73.0% for facial video with random spliced sound. Table 9 lists the average accuracy for each communication channel, on the basis of the PONS test, as well as significant differences among groups using Tukey's least significant difference (LSD) post hoc test for repeated measures (Rice, 1995). The cross-cultural accuracy of

almost every channel is different from the others. Compared with judgments of the voice alone, the addition of video information—especially the face—was associated with a large additional increase in cross-cultural accuracy. On average, the accuracy of judging the video channels did not benefit from the addition of content-filtered sound, though it did benefit from the addition of random-spliced sound.

Turning to the in-group advantage in accuracy based on the PONS test, the omnibus *F* test for nonverbal channel was also significant,  $F(10, 460) = 5.2$ ,  $p < 10^{-6}$ , suggesting that the in-group advantage varies according to the nonverbal channel. Table 9 lists the average in-group advantage in accuracy for each communication channel, based on the PONS, as well as significant differences among groups using Tukey's LSD post hoc test for repeated measures (Rice, 1995). Generally, it appears that the addition of sound to a silent channel reduces the in-group advantage. This suggests that providing information from additional channels of communication can reduce such cross-cultural differences.

### Specific Emotions

The studies in this meta-analysis included many different emotions, although there were only seven discrete emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise) and one global dimension (positivity–negativity) that were used in more than a few studies.

Many of the studies included in the meta-analysis reported data separately by emotion. This allowed us to calculate the percentage accuracy effect size for both cross-cultural accuracy and the in-group advantage for each emotion in each study. Table 10 summarizes these values. From these data, we can see that each

Table 6  
*Stem and Leaf Plot of Proportion In-Group Advantage in Cross-Cultural Emotion Recognition (N = 168)*

Stem	Leaf
.7	7
.5	3
.5	1
.4	
.4	6
.4	5
.4	
.4	1
.3	8, 9
.3	6
.3	5
.3	3
.3	0, 0, 1
.2	8, 9
.2	7, 7, 7
.2	
.2	2, 2, 3
.2	0, 1
.1	8, 8, 8, 9, 9, 9, 9
.1	6, 6, 6, 6, 7, 7
.1	4, 4, 4, 5, 5, 5, 5, 5, 5
.1	2, 2, 2, 2, 3, 3
.1	0, 0, 0, 0, 0, 0, 0, 1, 1, 1
.0	8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 9
.0	6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7
.0	4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5
.0	2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3
.0	0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1
-.0	0, 0, 1, 1, 1, 1
-.0	3, 3, 3
-.0	4, 4, 5, 5, 5
-.0	6, 6
-.0	8
-.1	0
-.1	2, 2, 3, 3
-.1	5
-.1	
-.2	0
-.2	2
-.2	5
-.2	
-.2	
-.3	0

emotion tested has both significant cross-cultural accuracy as well as a significant in-group advantage.

We found it an interesting bearing on the controversy regarding the status of contempt as a universal emotion (e.g., Izard & Haynes, 1988; Matsumoto, 1992b; Ricci Bitti et al., 1989; Russell, 1991) that although contempt showed a relatively low degree of in-group advantage (8.0%), it was also the most poorly recognized emotion (43.2%) cross-culturally.

To make comparisons across emotions, we performed an additional analysis using a repeated-measures ANOVA for both cross-cultural accuracy and the in-group advantage. This included those studies that reported data separately for the six most common emotions tested: anger, disgust, fear, happiness, sadness, and surprise. Contempt and positivity–negativity were, we regret, not

included along with the other six emotions frequently enough for us to include them in this analysis. The degree of cross-cultural accuracy in emotion recognition does appear to differ according to the particular emotion tested,  $F(5, 320) = 10.1, p < 10^{-14}$ . Fear and disgust were the most poorly recognized emotions, whereas happiness was the most accurately recognized. The degree of an in-group advantage in emotion recognition also appears to vary depending on the particular emotion tested. The omnibus  $F$  test across emotions was significant,  $F(5, 325) = 3.25, p < .008$ . The in-group advantage was lowest with happiness and anger, whereas it was greatest with fear and disgust.

*Differences in Emotions Across Channels*

The studies in this meta-analysis that reported data separately by emotion made use of several different channels of expression. In particular, data were sufficient to examine separately by emotion at the channels of facial photographs and voice, as summarized in Table 11 for cross-cultural accuracy and the in-group advantage. It is interesting that happiness was most accurately recognized from the face but least accurately recognized from the voice. In comparison, among those emotions with data for both channels, anger and sadness were most accurately recognized cross-culturally from the voice but relatively less accurately recognized from the face. Likewise, the in-group advantage was smallest for happiness in the face, but the highest for happiness in the voice. Fear showed a relatively large in-group advantage in both the face and the voice, whereas anger showed a relatively small in-group advantage in both the face and the voice.

*Cross-Cultural Exposure*

Is increased exposure to members of another culture associated with decreased in-group advantage in recognizing their emotional displays? We operationalized cross-cultural exposure using three different methods.

First, we compared the in-group advantage for those groups living together within the same nation, versus those groups that are separated by a national border. The rationale was that groups living together in the same country are likely to have greater exposure to one another. Results summarized at the top of Table 12 suggest that the degree of in-group advantage was smaller for such groups with greater exposure to each other,  $F(1, 164) = 8.6, p < .004$  ( $n = 120$  across nations,  $n = 46$  within nations). However, cross-cultural accuracy was no greater for groups living together within the same nation versus those separated by a national border ( $n = 45$  and 115, respectively),  $F(1, 158) = 0.3, p > .5$ .

Our second and third measures of cross-cultural exposure were proxies for cross-cultural communication. The physical proximity between groups was included under the rationale that groups living farther away are likely to have fewer opportunities to gain exposure to one another. After all, the bulk of studies included in the meta-analysis were conducted before advances in cross-cultural communication associated with the development of the Internet and cable television. Given that we operationalized physical distance as the as the crow flies distance between the capital cities, we were able to include only those cultural comparisons crossing a national border. A more direct proxy for cross-cultural communi-

Table 7  
Summary of Percentage Accuracy for Studies of Caucasian and East Asian Groups (N = 18)

Study	Stimulus group	In-group	Out-group	Caucasian accuracy (%)	East Asian accuracy (%)	In-group advantage (%)
<b>Caucasian stimuli</b>						
Beier & Zautra (1972)	U.S.	U.S.	Japan	52.0	37.6	14.4
Chan (1985)	U.S.	U.S.	Hong Kong	81.1	65.8	15.3
Ekman (1972), Study 4	U.S.	U.S.	Japan	83.0	74.1	8.9
Ekman et al (1987)	U.S.	U.S.	Hong Kong	86.4	79.8	6.6
	U.S.	U.S.	Japan	86.4	73.4	13.0
Izard (1971), Study 1	U.S.	U.S.	Japan	81.1	60.4	20.6
Izard (1971), Study 2	U.S.	U.S.	Japan	71.0	58.5	12.5
Leung & Singh (1998)	U.S.	U.S.	Hong Kong	89.2	72.9	16.3
Rosenthal et al. (1979)	North America	North America	Hong Kong	59.1	53.8	5.3
Russell et al. (1993)	North America	North America	Japan	62.3	47.3	15.0
Sweeney et al. (1980)	U.S.	U.S.	Japan	30.4	21.9	8.5
Van Bezooijen et al. (1983)	Netherlands	Netherlands	Taiwan	62.0	30.2	31.8
	Netherlands	Netherlands	Japan	62.0	26.1	35.9
Yik & Russell (1999)	North America	North America	Hong Kong	64.0	52.9	11.1
	North America	North America	Japan	64.0	56.3	7.7
Average				68.9	54.1	14.9
<b>East Asian stimuli</b>						
Markham & Wang (1996)	China	China	Australia	63.4	82.6	19.2
Sogon & Masutani (1989)	Japan	Japan	U.S.	49.7	53.6	3.9
Yik et al. (1998)	China	China	Canada	42.6	38.2	-4.5
Average				51.9	58.1	6.2

was the level of telephone communication between groups. This was the number of minutes per year of telephone traffic between the in-group and out-group nations, scaled by dividing by the number of telephones in the decoder's nation. The two measures were weakly correlated ( $r = .27, p < .002$ ), therefore they were included together into a single least squares multiple regression. The results, summarized in Table 13, reveal that the in-group advantage was smaller for groups that have greater exposure to each other. Heterogeneity analysis of the residuals, for those studies reporting the number of participants, suggests that no significant variance in the in-group advantage remains to be explained after accounting for variation in cross-cultural exposure,  $\chi^2(119) = 115.5, p > .5$ . Cross-cultural exposure was not signif-

icantly related to the universal recognition level of emotion across cultures,  $F(2, 112) = 1.3, p > .2$ .

*Minority Versus Majority Status*

Minority or majority group status, which was coded for cultural groups that live together within a single nation, was associated with the in-group advantage in emotion recognition. In particular, as summarized at the bottom of Table 12, minority group members were better able to judge the emotions of majority group members ( $n = 28$ ) than majority group members were, in return, able to judge the emotions of minority group members ( $n = 18$ ),  $F(1, 44) = 6.3, r = .35, p < .02$ . However, overall cross-cultural accuracy did not differ across minority ( $n = 27$ ) versus majority ( $n = 18$ ) groups,  $F(1, 43) = 2.4, p > .1$ .

To examine the pattern further, we conducted an additional analysis that controlled for extraneous sources of variance. In the previous comparison, we included all studies for which minority or majority status could be coded. In the next analysis, we examined only those balanced studies that included both majority groups judging minority groups and minority groups judging majority groups. This was done to reduce the error that might be associated with variance in experimental methods and task difficulty. Consequently, this second analysis showed a larger effect size,  $t(10) = 1.8, r = .50, p < .10$ , although with 11 such studies the effect reached only marginal significance. Table 14 summarizes the studies included in this second analysis. Given the stability of the direction of the in-group advantage elsewhere in this meta-analysis, it is worth noting that this advantage reverses direction in 7 of the 11 studies listed.

*Quality and Validity of Studies and Other Moderators*

Do the current results differ depending on the quality and validity of the methods used by individual researchers? Many

Table 8  
Stem and Leaf Plot of Mean Effect Size (r) for In-Group Advantage in Cross-Cultural Emotion Recognition (N = 48)

Stem	Leaf
.8	7
.7	6
.6	9
.5	7, 9
.4	7, 7, 7, 9, 9
.3	0, 2, 3, 5, 5, 6, 8, 9
.2	1, 1, 1, 2, 2, 6, 6, 6, 7
.1	3, 3, 5, 9
.0	0, 0, 0, 0, 2, 2, 5, 7, 7, 9
-.0	4, 6, 6
-.1	0, 3, 5
-.2	
-.3	
-.4	
-.5	
-.6	3

Table 9  
*Cross-Cultural Accuracy and In-Group Advantage in Emotion Recognition Across Nonverbal Channels*  
 Based on Data From the PONS Test ( $N = 47$ )

Nonverbal channel	Cross-cultural accuracy (%)					In-group advantage (%)				
	No video	Body	Figure	Face	<i>M</i>	No video	Body	Figure	Face	<i>M</i>
No sound ( <i>SD</i> )		53.6 <sub>d</sub> (10.0)	56.7 <sub>e</sub> (12.4)	58.7 <sub>e</sub> (8.1)	56.3		6.8 <sub>b,c</sub> (9.8)	10.6 <sub>d</sub> (11.6)	7.9 <sub>c,d</sub> (7.7)	8.4
Random spliced ( <i>SD</i> )	25.2 <sub>b</sub> (7.8)	52.4 <sub>d</sub> (7.1)	67.2 <sub>g</sub> (7.6)	73.0 <sub>h</sub> (10.1)	54.4	5.0 <sub>a,b</sub> (9.9)	4.7 <sub>a,b</sub> (7.9)	4.4 <sub>a</sub> (7.0)	6.9 <sub>b,c</sub> (9.9)	5.2
Content filtered ( <i>SD</i> )	19.6 <sub>a</sub> (7.4)	42.7 <sub>c</sub> (9.2)	68.2 <sub>g</sub> (9.4)	61.5 <sub>f</sub> (8.1)	48.0	5.1 <sub>a,b</sub> (8.3)	8.2 <sub>c,d</sub> (10.5)	5.7 <sub>a,b,c</sub> (8.6)	6.3 <sub>a,b,c</sub> (8.3)	6.3
<i>M</i>	22.4	49.6	64.0	64.4		5.0	6.6	6.9	7.0	

*Note.* Differences between means that do not have a subscript in common are significant at the .05 level or beyond by repeated-measure least significant difference post hoc contrasts. Subscripts are alphabetized according to ascending numerical order. PONS = Profile of Nonverbal Sensitivity (Rosenthal et al., 1979).

researchers who study emotional communication have detailed what features they believe are required for a study to be valid (e.g., Matsumoto, 1992a; Rosenthal et al., 1979; Russell, 1993, 1994; Ekman, 1994; Izard, 1994). However, these criteria are not the same across all researchers in the field, and have been the subject of ongoing debate (e.g., Ekman, 1994; Izard, 1994; Russell, 1994). Therefore, to avoid imposing our own opinion of what makes a study valid, we followed Glass's (1978) reasoning regarding quality coding in meta-analyses. "The sensible course to follow," he wrote, "is to describe—in quantitative terms—features of designs and correlate them with the study findings: The obtained relationships will reveal how important matters of design are and precisely what to do about them" (p. 3). Accordingly, we coded the current studies for objective features of the study that could contribute to the validity or quality of a study. Each of these features differed across research teams and had the potential to affect quality or validity. These features are whether the study had a balanced design—that is, whether every cultural group in the study judged emotions expressed by members of every other group; whether the recognizability of stimuli was validated prior to the study through the use of a separate consensus sample of raters; whether the emotions in the stimuli were expressed spontaneously, were posed, or were imitated; whether participants responded using multiple-choice, dimensional, or free-response scales; whether the socio-

economic status of the cultural groups was matched; and finally, whether the study was published or unpublished.

*Research team.* Before examining these dimensions of quality, we first investigated whether cross-cultural accuracy or the in-group advantage in emotion recognition differed across research teams. Given that different groups tend to make different design choices, this is a first indication of whether study quality is an important moderator of the current results. Cross-cultural accuracy of emotion recognition differed significantly across research teams;  $F(4, 112) = 8.2, p < 10^{-5}$ , for studies crossing national borders; and,  $F(3, 43) = 3.0, p < .04$ , for studies within a single nation, as summarized in Table 15. In particular, studies conducted by Ekman and Matsumoto, Izard, Scherer, and Nowicki were associated with greater cross-cultural accuracy than were studies conducted by other researchers. However, the extent of the in-group advantage did not differ significantly across research teams;  $F(4, 117) = 1.6, p > .1$ , across nations; and,  $F(3, 43) = 0.1, p > .9$ , within nations.

*Balanced design.* In studies with balanced designs, members of every cultural group in the study judged emotions expressed by members of every other group. This type of design is optimal because it allows for separate analysis of effects across recognizability of emotional expressions, skill in judging emotional expressions, and the interaction between such encoding and decoding

Table 10  
*Summary of Cross-Cultural Accuracy and In-Group Advantage in Emotion Recognition Across Individual Emotions*

Emotion	Cross-cultural accuracy						In-group advantage					
	<i>M</i> (%)	<i>SD</i> (%)	95% CI	<i>p</i>	<i>r</i>	<i>N</i>	<i>M</i> (%)	<i>SD</i> (%)	95% CI	<i>p</i>	<i>r</i>	<i>N</i>
Overall	58.0	19.6	54.9–61.0	<.00001	.95	162	9.3	14.2	7.2–11.5	<.00001	.55	168
Anger	64.9	21.3	60.2–69.6	<.00001	.95	83	7.7	13.9	4.7–10.7	<.00001	.49	83
Contempt	43.2	31.7	32.8–53.6	<.00001	.81	38	8.0	18.9	1.4–14.6	.00955	.39	34
Disgust	60.6	27.0	54.0–67.1	<.00001	.91	67	15.1	22.5	9.7–20.6	<.00001	.56	67
Fear	57.5	24.5	52.1–62.8	<.00001	.92	82	13.1	22.8	8.1–18.1	<.00001	.50	82
Happiness	79.1	26.7	73.2–85.0	<.00001	.95	81	7.3	16.0	3.7–10.8	.00005	.41	81
Positive–negative	54.1	10.4	48.9–59.2	<.00001	.98	18	5.3	4.4	3.3–7.3	.00001	.77	22
Sadness	67.5	19.5	63.2–71.9	<.00001	.96	81	10.2	18.2	6.2–14.2	<.00001	.49	81
Surprise	67.6	27.2	61.1–74.1	<.00001	.93	70	11.8	19.9	7.0–16.5	<.00001	.51	70

*Note.* It was not possible to separate results by individual emotion in all studies. CI = confidence interval.

Table 11  
*Cross-Cultural Accuracy and In-Group Advantage in Emotion Recognition by Particular Emotion and Channel*

Emotion	Facial photographs				Voice			
	<i>N</i>	<i>M</i> (%)	<i>SD</i> (%)	95% CI	<i>N</i>	Average (%)	<i>SD</i> (%)	95% CI
Cross-cultural accuracy								
Anger	70	65.2 <sub>b,c</sub>	22.4	59.9–70.4	12	63.7 <sub>c</sub>	14.9	55.2–72.2
Contempt	34	46.9 <sub>e</sub>	31.3	36.4–57.4		—		
Disgust	63	62.1 <sub>c,d</sub>	26.6	55.5–68.7		—		
Fear	70	58.3 <sub>d</sub>	25.2	52.4–64.2	11	50.8 <sub>a,b</sub>	20.4	38.8–62.9
Happiness	69	87.6 <sub>a</sub>	17.1	83.5–91.6	11	28.9 <sub>a</sub>	18.6	17.9–39.8
Sadness	68	68.4 <sub>b</sub>	20.7	63.5–73.3	12	62.8 <sub>b,c</sub>	11.4	56.4–69.3
Surprise	66	69.1 <sub>b</sub>	27.0	62.6–75.6		—		
In-group advantage								
Anger	70	7.8 <sub>b,c</sub>	14.1	4.5–11.2	12	7.4 <sub>a</sub>	14.0	–0.6–15.3
Contempt	30	4.4 <sub>d</sub>	16.3	–1.4–10.3		—		
Disgust	63	15.0 <sub>a</sub>	22.9	9.3–20.6		—		
Fear	70	13.4 <sub>a,b</sub>	23.9	7.8–19.1	11	12.9 <sub>a</sub>	14.7	4.2–21.6
Happiness	69	5.8 <sub>c,d</sub>	14.2	2.4–9.1	11	17.5 <sub>a</sub>	23.2	3.8–31.3
Sadness	68	10.7 <sub>a,b</sub>	18.9	6.2–15.2	12	7.6 <sub>a</sub>	15.4	–1.2–16.3
Surprise	66	11.2 <sub>a,b</sub>	20.1	6.3–16.0		—		

*Note.* Differences between means that do not have a subscript in common are significant at the .05 level or beyond by least significant difference post hoc contrasts. Subscripts are alphabetized according to ascending numerical order. Categories are not reported if there were three or fewer such studies (represented by dashes).

skills. However, balanced designs are more difficult to conduct, and therefore are less common in the present sample. The size of the in-group advantage did not differ significantly across balanced versus unbalanced designs,  $F(1, 119) = 3.2, p = .08$ , for  $n = 14$  balanced and  $n = 107$  unbalanced samples crossing national borders;  $F(1, 46) = 0.2, p > .6$ , for  $n = 25$  balanced and  $n = 23$  unbalanced samples within a single nation. However, cross-cultural accuracy was significantly lower in balanced versus other designs,  $F(1, 114) = 15.6, p < .0002$ , for  $n = 14$  balanced and  $n = 102$  unbalanced samples across nations;  $F(1, 45) = 7.3, p < .01$ , for  $n = 25$  balanced and  $n = 22$  unbalanced samples within nations.

*Consensus validation of stimuli.* Some researchers used all of the emotional stimuli that they collected, whereas others attempted to remove examples of emotional expressions that were idiosyncratic or difficult to recognize by including a separate consensus sample of raters judging the recognizability of stimuli. The size of the in-group advantage did not significantly differ across studies with stimuli recognition validated through a consensus sample,  $F(1, 109) = 0.3, p > .5$ , for  $n = 87$  consensus validated and  $n = 24$  nonconsensus validated samples crossing national borders,  $F(1, 46) = 1.1, p > .3$ , for  $n = 29$  consensus validated and  $n = 19$  nonconsensus validated samples within a single nation. However, cross-cultural recognition accuracy was higher for studies with stimuli recognition levels that had been validated through consensus,  $F(1, 104) = 10.3, p < .002$ , for  $n = 86$  consensus validated and  $n = 20$  nonconsensus validated samples across nations;  $F(1, 45) = 5.3, p < .03$ ,  $n = 28$  consensus validated and  $n = 19$  nonconsensus validated samples within nations.

*Manner of expressing emotional stimuli.* Studies elicited emotional stimuli through the use of several different methods. Some studies made use of spontaneous emotions and validated which

emotion was expressed through a criterion such as the antecedent context eliciting the emotion. Others asked participants to pose emotions or to imitate emotional expressions that had been chosen on the basis of a priori theoretical grounds. Table 16 presents the results of a categorical analysis that examines whether the findings differ across studies on the basis of the method used for eliciting emotional expressions. The extent of the in-group advantage did not differ significantly on the basis of the method used for eliciting emotional expressions,  $F(2, 119) = 1.4, p > .2$ , for studies crossing national borders;  $F(1, 46) = 0.0, p > .8$ , for studies within a single nation. Cross-cultural accuracy differed significantly across methods,  $F(2, 114) = 20.1, p < 10^{-7}$ , across nations;  $F(1, 45) = 10.8, p < .002$ , within nations, such that imitated emotions were recognized more accurately than either spontaneous or posed emotions.

*Response format.* Most studies made use of multiple-choice response formats, but some studies also used dimensional scales (Ricci Bitti, Giovannini, Argyle, & Graham, 1979; Yik, Meng, & Russell, 1998), or free responses by participants (Gates, 1923; Haidt & Keltner, 1999; Kellogg & Eagleson, 1931; and Yik et al., 1998). Because so few studies used response types other than multiple choices—and because of the seven samples contained within these six studies, three of these made use of more than one type of response format in the same study—it was not possible to perform an ANOVA across the response-type categories. However, it is worth noting that the cross-cultural percentage accuracy values attained in these six studies (see entries in Table 1) are all well in excess of zero, which provides evidence for substantial cross-cultural accuracy in examining emotion recognition using response formats other than multiple choice.

*Matching for socioeconomic status.* We tested whether discrepancies in socioeconomic status among members of ethnic

Table 12  
*Categorical Model for the Effect of Type of Cultural Group on the In-Group Advantage in Emotion Recognition*

Effect	n	In-group advantage		Within-class heterogeneity
		Average effect size (%)	95% CI	
Type of cultural group		$F(1, 164) = 8.6, p < .004$		
Across nations	120	11.0	8.6–13.4	71.1
Within a nation	46	4.1	0.0–8.2	22.1
Majority versus minority status		$F(1, 44) = 6.3, p < .02$		
Majority group judging minority	18	10.1	3.3–16.8	16.5
Minority group judging majority	28	0.2	–4.7–5.1	12.4

*Note.* Two samples containing both across-nation and within-nation analyses combined were excluded from this analysis. CI = confidence interval.

groups would be associated with emotion recognition accuracy or the in-group advantage. In studies without such discrepancies ( $n = 24$  out of 46 samples measuring cross-cultural accuracy and 47 samples measuring the in-group advantage, respectively), either (a) participant groups had matched levels of social class or (b) social class was measured and its effect was partialled out of the cross-ethnic group analysis. There was no relationship between socioeconomic discrepancy and either cross-cultural emotion recognition accuracy,  $F(1, 44) = 1.1, p > .2$ , or the magnitude of the in-group advantage,  $F(1, 45) = 0.3, p > .5$ .

*Publication status and file drawer effects.* We tested whether the magnitude of cross-cultural accuracy and the in-group advantage in emotion recognition was related to publication status. Published works included journal articles, books, book chapters, and manuscripts accepted for publication. Unpublished works included doctoral and master's-level theses, papers presented at conferences that were not included in the conference proceedings, and manuscripts not submitted or not yet accepted for publication. The magnitude of the in-group advantage did not differ signifi-

cantly across published versus unpublished studies,  $F(1, 120) = 1.6, p > .2$ , for  $n = 114$  published and  $n = 8$  unpublished samples crossing national borders,  $F(1, 46) = 0.0, p > .9$ , for  $n = 33$  published and  $n = 15$  unpublished samples within a single nation. For studies including cultural groups across nations, published studies ( $n = 113$ ) documented higher cross-cultural accuracy in emotion recognition than the few ( $n = 4$ ) such unpublished studies,  $F(1, 115) = 9.6, p < .003$ . By contrast, cross-cultural accuracy did not differ between published ( $n = 31$ ) and unpublished ( $n = 15$ ) studies conducted within a single nation,  $F(1, 44) = 2.6, p > .1$ .

Many critics of meta-analysis have argued that biases in the publication criteria of editors are reflected in biased samples of articles used in reviews. We attempted to address this “file drawer problem” (Rosenthal, 1991, p. 128) in three different ways. First, we attempted to solicit manuscripts from researchers’ file drawers, as described above. Second, we compared the results of the published versus unpublished studies. Finally, we computed a sensitivity analysis designed to measure just how many items must languish in file drawers before the results of this meta-analysis would be affected. Using Rosenthal’s (1991) formulas—along with his suggestion that truncating significance levels to a minimum of .05 would provide a highly conservative test—we found that it would take at least 771 studies with an average effect size ( $r$ ) of 0 and a  $p$  value of .5 for the in-group advantage to become just barely significant at .05 (one-tailed). Note that this value is an underestimate because it is based only on those ( $n = 48$ ) studies for which an effect size could be calculated. That is, to make the present results just barely significant would require more than 16 times more null studies than the 48 original studies for which an independent significance level could be calculated.

A similar analysis is possible for the percentage accuracy effect size, although some modification is necessary. The standard formulas were designed for meta-analyses of studies that reported individual significance levels, which is not the case for the percentage accuracy analyses. Instead, we performed a sensitivity analysis in which we examined the results of adding hypothetical studies with negative results. In the case of the cross-cultural data, we would be able to add 802 additional studies with percentage accuracy of  $-10\%$ —smaller than the smallest value actually included in the meta-analysis—and still have an overall significance

Table 13  
*Continuous Model for the Effect of Cross-Cultural Exposure on the In-Group Advantage in Emotion Recognition*

Predictor	In-group advantage		
	$b(\%)$	95% CI for $b$	$\beta$
Physical proximity <sup>a</sup>	–0.6**	–1.3–0.0	–0.18**
Telephone communication <sup>b</sup>	–0.6**	–1.2–0.0	–0.18**
Additive constant	8.3**	0.9–15.8	

*Note.* This model is a least squares multiple regression. Overall  $F$  test of model:  $F(2, 119) = 5.6, p < .005$ ; heterogeneity of residuals:  $\chi^2(119) = 115.5, p > .5$ ; Multiple  $R = .29$ ;  $b$  = unstandardized regression coefficient; CI = confidence interval;  $\beta$  = standardized regression coefficient.

<sup>a</sup> Physical proximity is per 1,000 kilometers distance between capital cities (reverse coded). Negative value indicates that greater exposure is associated with lower degree of in-group advantage. <sup>b</sup> Telephone communication is number of minutes per year telephone traffic between the in-group and out-group nations, scaled by the number of telephones in the decoder’s nation (square-root transformed). Negative value indicates that greater telephone communication is associated with lower degree of in-group advantage.

\*\*  $p < .05$ , two-tailed.



Table 14  
*Relationship Between Majority or Minority Group Status and In-Group Advantage in Emotion Recognition*

Study	Majority judges minority (%)	Minority judges majority (%)	Difference (%)
Albas et al. (1976)	29.4	27.2	2.2
Buchman (1973)	-3.8	0.6	-4.4
Gitter et al. (1972a)	-1.4	-5.8	4.4
Gitter et al. (1972b)	5.2	-4.1	9.3
Gitter et al. (1971)	-13.3	5.3	-18.7
Mehta et al. (1992)	-12.4	20.6	-33.0
Nowicki et al. (1998), Study 1	36.7	-22.8	59.4
Nowicki et al. (1998), Study 2	30.0	-20.6	50.6
Ricci Bitti et al. (1979)	22.3	-30.5	52.8
Ricci Bitti et al. (1980)	21.3	-25.3	46.7
Wolfgang (1980)	10.5	-8.2	18.7

level for the predicted effect of less than .05 (one-tailed). In the case of the in-group advantage, we would be able to add 177 additional studies with a percentage accuracy of -6.6%—the value at the 7.5th percentile of the studies included in the meta-analysis—and still have an overall significance level for the predicted effect of  $p < .05$  (one-tailed). It would be unlikely to find this many null studies in file drawers, and thus the current meta-analysis appeared to be robust against the file drawer problem.

*Other moderating factors.* We examined the effects of several other possible moderators associated with emotion recognition across cultures, in particular the year of the study, the age of participants, and the gender of participants. In continuous models examining all studies for which these variables could be coded, none significantly moderated the size of the in-group advantage in emotion recognition,  $F(3, 64) = 0.6, p > .6$ , for studies crossing national borders;  $F(3, 30) = 1.0, p > .4$ , for studies within a single nation. Among studies conducted across national borders, cross-cultural accuracy appeared to be greater for more recent studies

( $b = 0.7\%$ , 95% CI = 0.2% to 1.2%,  $p < .01$ ), whereas no variables significantly moderated cross-cultural accuracy in the pool of studies conducted within a single nation,  $F(3, 30) = 1.8, p > .1$ .

*Model including multiple moderators.* To examine the effect of multiple moderators simultaneously, we conducted a multiple regression model, jointly examining all variables that proved significant on a univariate basis. We dummy coded categorical variables so that all variables could be included in the same analysis. Given that cross-cultural exposure was the only significant univariate predictor of the magnitude of the in-group advantage, this multiple-moderators analysis examined only cross-cultural accuracy in emotion recognition.

Intercorrelations among variables revealed that the dummy codes for research teams related closely to other significant moderators in the current analysis, as illustrated in Table 17. In particular, some research groups were more likely to use certain nonverbal channels, to use certain methods of posing stimuli, to use balanced designs, to use consensus samples, and to conduct work during earlier or later years. Therefore, such variables were used in the multivariate models in lieu of research teams.

For cross-cultural accuracy in emotion recognition, a single regression model included both across-nation and within-nation studies. Results presented above indicated no significant difference between the two types, and only one individual moderator showed a different effect in both types. Unpublished status of a manuscript was associated with lower accuracy in the few such studies across nations, whereas it was associated with marginally greater accuracy in studies within a nation. Considering both across-nation and within-nation studies together, the effect of publication status was no longer significant,  $F(1, 159) = 0.6, p > .4$ . Therefore, the variable was excluded from the multivariate model.

All significant univariate moderators of cross-cultural accuracy were included in the multivariate model after examination of intercorrelation as illustrated in Table 18. From this table, we can see that use of a balanced design strongly negatively correlated with both the use of a consensus sample and imitation as a method of eliciting emotional stimuli—these design choices appeared to

Table 15  
*Categorical Model for the Effect of Research Team and Stimuli on Emotion Recognition*

Cultural group and research team	Cross-cultural accuracy			In-group advantage			
	<i>n</i>	Average effect size (%)	95% CI	<i>n</i>	Average effect size (%)	95% CI	Within-class heterogeneity
Across nations		$F(4, 112) = 8.2, p < 1 \times 10^{-5}$			$F(4, 117) = 1.6, p < .19$		
Ekman & Matsumoto	54	65.9 <sub>b</sub>	59.8–72.1	55	14.8 <sub>a</sub>	10.3–19.3	77.8**
Izard	12	67.7 <sub>b</sub>	60.4–74.9	12	11.7 <sub>a</sub>	3.7–19.7	4.0
Rosenthal	10	52.0 <sub>a</sub>	48.5–55.6	10	6.8 <sub>a</sub>	3.6–10.0	2.6
Scherer	9	56.5 <sub>a,b</sub>	51.2–61.9	9	10.7 <sub>a</sub>	5.4–16.1	1.4
Other researchers	32	43.4 <sub>a</sub>	36.8–50.1	36	8.1 <sub>a</sub>	3.7–12.5	43.8*
Within a nation		$F(3, 43) = 3.0, p < .04$			$F(3, 43) = 0.1, p > .9$		
Ekman & Matsumoto	7	70.1 <sub>b</sub>	43.0–97.1	7	5.5 <sub>a</sub>	-11.7–22.7	8.6
Rosenthal	4	49.8 <sub>a,b</sub>	44.4–55.2	4	6.5 <sub>a</sub>	0.8–12.3	0.1
Nowicki	7	64.2 <sub>a,b</sub>	50.3–78.1	7	6.4 <sub>a</sub>	-14.5–27.3	16.8***
Other researchers	29	51.5 <sub>a</sub>	45.9–57.0	29	3.8 <sub>a</sub>	-1.5–9.0	9.4

*Note.* Values were obtained using omnibus *F* tests. Differences between means that do not have a subscript in common are significant at the .05 level or beyond by least significant difference post hoc contrasts.

\*  $p < .10$ . \*\*  $p < .05$ . \*\*\*  $p < .01$ .

Table 16  
*Categorical Model for the Effect of Manner of Expressing Emotional Stimuli on Emotion Recognition*

Cultural group and manner of expression	Cross-cultural accuracy			In-group advantage			Within-class heterogeneity
	<i>n</i>	Average effect size (%)	95% CI	<i>n</i>	Average effect size (%)	95% CI	
Across nations	$F(2, 114) = 20.1, p < 1 \times 10^{-7}$			$F(2, 119) = 1.4, p < .26$			
Spontaneous	4	39.5 <sub>a</sub>	26.0–53.0	8	3.8 <sub>a</sub>	–5.0–12.5	2.7
Posed	45	46.0 <sub>a</sub>	40.4–51.7	45	11.3 <sub>a</sub>	7.1–15.6	65.9**
Imitated	68	67.1 <sub>b</sub>	62.7–71.5	69	12.6 <sub>a</sub>	9.1–16.1	71.3
Within a nation	$F(1, 45) = 10.8, p < .002$			$F(1, 46) = 0.0, p > .8$			
Spontaneous	0			0			
Posed	32	50.6 <sub>a</sub>	44.9–56.3	32	4.4 <sub>a</sub>	–2.0–10.7	33.5
Imitated	15	67.5 <sub>b</sub>	57.8–77.2	16	5.0 <sub>a</sub>	0.9–9.0	2.7

Note. Values were obtained using omnibus *F* tests. Differences between means that do not have a subscript in common are significant at the .05 level or beyond by least significant difference post hoc contrasts.

\*\*  $p < .05$ .

represent a tradeoff. Further, studies using vocal tone were less likely to use imitation for stimuli. Over time, studies have become more likely to use consensus samples and multiple-choice response. Given these strong intercorrelations, the results of the multiple regression model must be interpreted with caution. Table 19 summarizes the results of this multivariate ordinary least squares regression. In this model, cross-cultural accuracy was greater for emotional stimuli elicited through imitation as well as chronologically later studies, and was marginally lower for studies using a balanced design. Although this model accounted for an adjusted  $R^2 = 30.5\%$  of the variance associated with cross-cultural accuracy, a heterogeneity analysis of the residuals on those studies for which the number of participants was reported,  $\chi^2(147) = 220.1, p < .0001$ , revealed that significant variation still remained unexplained.

### Discussion

The results of this work support an interactionist interpretation of emotion recognition. Evidence for the cross-cultural recognition of emotions in the current meta-analysis suggests that certain core components of emotions are universal and likely biological. These analyses also present evidence, however, that emotional expressions may lose some of their meaning across cultural boundaries. We found evidence that emotions may be more accurately understood when they are judged by members of the same national,

ethnic, or regional group that had expressed the emotion. This in-group advantage indicates that culture can have an important role in shaping our emotional communication.

### *Evaluating Explanations for Cultural Variability in Emotion Recognition*

The in-group advantage—whereby we understand emotions more accurately when they are expressed by members of our own cultural or subcultural group—is one of several possible explanations that has been proposed to account for cultural variability in emotion recognition. The present meta-analysis provides support for this explanation. The results of this meta-analysis suggest that theories of display rules and decoding rules do not provide a complete explanation regarding cultural variability in emotion recognition rates. It is interesting that whereas proponents of display rule theory (e.g., Ekman, 1972) argue that the stimuli used in Ekman's studies are carefully pretested to exclude display rules, the meta-analytic results indicate that studies using these stimuli seem to show no smaller a degree of in-group advantage versus other studies. Although it is possible that display rules could lead individuals to express their emotions more accurately to members of their own cultural group, all of the studies included in the analysis used prerecorded emotional expressions. The results of the current meta-analysis suggest that the match between the cultural background of the expressor and judge is important. The-

Table 17  
*Correlations Between Research Teams and Study Methodology*

Study methodology	<i>n</i>	Research team				
		Ekman & Matsumoto	Nowicki	Izard	Rosenthal	Scherer
Consensus sample	157	.11	.14*	.18*	.19*	.15*
Stimuli elicited through imitation	168	.59§	–.23***	.29****	–.31§	–.24***
Balanced design	168	–.33§	.15*	–.16**	–.16**	–.13*
Multiple choice	168	.15*	.04	.06	.06	.05
Year of study	168	.15*	.28§	–.23**	–.05	.30§
Vocal channel	168	–.50§	.04	–.19**	.47§	.37§

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ . \*\*\*\* $p < .001$ . § $p < .0001$ , all two-tailed.

Table 18  
Correlations Among Significant Univariate Moderators of Cross-Cultural Emotion Recognition Accuracy

Moderators	1	2	3	4	5	6
1. Consensus sample	—					
2. Balanced design (N)	-.58§ (171)	—				
3. Stimuli elicited through imitation (N)	.32§ (170)	-.34§ (181)	—			
4. Vocal channel (N)	-.08 (157)	.12 (168)	-.65§ (168)	—		
5. Year of study (N)	.31§ (157)	-.14* (168)	.11 (168)	.00 (168)	—	
6. Multiple choice (N)	.02 (160)	-.04 (171)	.19** (171)	.12 (168)	.25*** (168)	—

\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ . § $p < .0001$ , all two-tailed.

orists (e.g., Matsumoto, 1989, 1992a) have noted that members of collectivistic cultures such as the Japanese may demonstrate decreased emotion recognition performance when judging stimuli from other regions, particularly negative emotions. However, they may be able to outperform their individualistic counterparts when judging stimuli originating in Japan (e.g., Hatta & Nachshon, 1988; Sogon & Masutani, 1989).

A number of other explanations that have been offered to explain cultural variability in accuracy rates concern the relationship between emotion recognition and differences in language (e.g., Matsumoto & Assar, 1992; Mesquita & Frijda, 1992; Mesquita et al. 1997). The data from the present meta-analysis are consistent with several of these explanations—for example, translation difficulties (Mesquita & Frijda, 1992; Mesquita et al. 1997) and distraction from expressive content due to baseline differences in voice quality across language families (Mesquita & Frijda, 1992). However, the present results are less consistent with other possible explanations related to language, such as differing suitability of emotional vocabulary (Matsumoto & Assar, 1992). It is notable that the in-group advantage was found consistently among groups that share the same native language. For example, we found cultural variability in the form of in-group advantage when English-speaking groups such as the English, Scottish, Irish, New Zealanders, and Australians judged the emotional expressions of Americans. These findings cast doubt on explanations that rely on individual differences in languages, such as that of Matsumoto and Assar (1992), who suggested that cultural differences in emotion recognition result from greater suitability of some languages versus others to express emotional concepts. However, these findings cannot resolve the frequent overlap between participants' language and their cultural group membership. It would be valuable to tease apart this confound for a more precise understanding of the relationship between language and cross-cultural differences in emotion recognition.

*Accounting for the In-Group Advantage*

Why would we better understand emotions expressed by a member of our own cultural, ethnic, or regional in-group? Several possible explanations are suggested by research in this field. In

particular, cultural learning and expressive style (Albas, et al., 1976; Allport & Vernon, 1933; Scherer et al., 2001), differences in emotional concepts (Russell & Yik, 1996), and cognitive representations (Anthony, Copper, & Mullen, 1992) may all play a role. In future work, it would be desirable to measure directly the variables that can distinguish among these possible explanations.

Recent research has documented the cultural specificity of emotional experience and linguistic expressions (Fiske et al., 1998; Harré, 1986; Markus & Kitayama, 1991; Mesquita & Frijda, 1992; Mesquita et al., 1997; Scherer & Wallbott, 1994). The culture-specific elements of emotional behavior must be learned, either by growing up in the culture or else by later exposure to the culture. The results of this meta-analysis are consistent with the theory of cultural learning in emotional behavior. First, individuals may be able to recognize more accurately emotions expressed by members of their own culture, which suggests the presence of culture-specific elements of emotional behavior. Furthermore, greater cultural exposure was also associated with a smaller in-group advantage.

There may be linguistic and conceptual reasons contributing to the possibility that people more accurately recognize in-group emotions. Cultural differences in emotional concepts lead to both methodological issues in the translation of words as well as theoretical issues in the translation of experience. The culture in which emotional concepts originate may be as important for predicting recognition accuracy as the culture in which the physical emotional expressions originate (Russell & Yik, 1996). Therefore, the in-group advantage may result from the match between target and judge in the origin of the emotional expressions as well as the origin of the emotional concepts.

Finally, given the research on facial recognition, it is possible that people use different and perhaps more efficient modes of processing when judging in-group members' expressions. The in-group advantage documented by this meta-analysis mirrors a similar finding in the area of face recognition, whereby same-race faces are identified more accurately and efficiently than different-race faces (O'Toole et al., 1996). The in-group advantage in facial recognition has been linked to differences in the level of processing for in-group versus out-group faces, whereby individuals more

Table 19  
Multiple Regression Model for Cross-Cultural Accuracy in Emotion Recognition

Predictor	<i>b</i> (%)	95% CI for <i>b</i>	$\beta$
Consensus sample	3.5	-4.4-11.5	0.08
Balanced design	-7.5*	-15.5-0.5	-0.17*
Stimuli elicited through imitation	12.0***	3.5-20.5	0.31***
Vocal channel	-2.4	-10.9-6.2	-0.06
Multiple choice	8.7	-7.2-24.6	0.09
Year of study	0.3**	0.1-0.5	0.18**
Additive constant	-530.4	-999.2--61.7	

Notes. This model is a least squares multiple regression. Overall *F* test of model:  $F(6, 144) = 10.5, p < 1 \times 10^{-8}$ . Heterogeneity of residuals:  $\chi^2(147) = 220.1, p < .0001$ . Multiple *R* = .55; *b* = unstandardized regression coefficient; CI = confidence interval;  $\beta$  = standardized regression coefficient.  
\* $p < .10$ . \*\* $p < .05$ . \*\*\* $p < .01$ , all two-tailed.

often process same-race facial stimuli in an exemplar mode that handles individual differences effectively (Anthony et al., 1992). A similar mechanism may exist for emotion recognition. Such a phenomenon could contribute to but could not completely explain the current findings. In many of the experiments, it is difficult to propose a mechanism by which cross-cultural participants would know that they were judging out-group members unless an identifier of group membership is contained in the emotional expression itself. For example, some studies used Caucasian photographs and multiple Caucasian participant groups or content-free tone-of-voice. This raises the intriguing possibility that subtle differences in expressive styles across cultural groups might serve as an identifier similar to that of race in facial recognition studies.

### *Minority–Majority Group Relations*

The ability of individuals from different cultural groups to understand one another's emotions is not always symmetric; minority group members showed higher accuracy when judging majority groups than majority group members showed in return when judging minority group members. This effect was often so large that minority groups actually understood the majority's emotional expressions better than they understood their own. In comparison with the relatively stable pattern of in-group advantage found elsewhere in this review, minority groups within a nation often showed an *out-group* advantage. This occurred in 7 of the 11 balanced-design studies (64%), compared with 35.7% of the larger group of within-nation studies. These results indicate that intergroup relations and differences in power and exposure across groups may be relevant to the in-group advantage in nonverbal accuracy.

### *Nonverbal Channels*

An additional goal in conducting the present meta-analysis was to explore whether universals and cultural differences in emotion varied across nonverbal channels of expression. We found preliminary evidence that this may be the case; more complex dynamic stimuli were associated with a marginally lower degree of in-group advantage than simpler static stimuli, but they were associated with lower cross-cultural accuracy as well. This suggests that posed photographs may be somewhat stylized and exaggerated to improve legibility, using conventions that may be partially culture-specific. By contrast, dynamic stimuli provide richer and more ecological information that might be more difficult to decode out of context, but may be less susceptible to cultural conventions.

It seems timely to call for greater attention to be paid to the richness and complexity of emotion in order to increase the ecological validity of cross-cultural research on emotion. One way to do so would be to investigate multiple nonverbal channels to assess emotional expression in the dynamic stream of behavior. As Bruner and Tagiuri noted back in 1954, we should move beyond the use of "the human face in a state of arrested animation" (p. 638). At that time, the use of other forms of stimuli were more troublesome, but today the consumer market for high-quality audio and video equipment makes it increasingly easy and inexpensive for researchers to heed Bruner and Tagiuri's call.

### *Individual Emotions*

Another goal of the meta-analysis was to explore the cross-cultural judgment of specific emotions. Results revealed that happiness and anger showed the smallest degree of an in-group advantage, as well as high cross-cultural recognition accuracy. Perhaps this reflects attunement across cultures to accurate recognition of signals of approach and avoidance (Baron & Boudreau, 1987; McArthur & Baron, 1983).

This meta-analysis provides preliminary evidence to inform two current debates in the literature concerning specific emotions, although differences in the emotions tested across studies prevented exhaustive analyses. One recent debate has centered on whether contempt should be included as one of the so-called basic emotions. (e.g., Izard & Haynes, 1988; Matsumoto, 1992b; Ricci Bitti et al., 1989; Russell, 1991) Results revealed that contempt did show significant cross-cultural accuracy, although it had the lowest accuracy level of the individual emotions. Second, there has been a debate about whether general emotional dimensions are more universally recognized than specific categories (e.g., Russell, 1994). In our analysis, we found that positivity–negativity was associated with the smallest degree of an in-group advantage, although it was also associated with relatively low cross-cultural accuracy, in part because it was tested primarily in studies with dynamic nonverbal channels.

### *Interaction of Nonverbal Channels and Emotions*

Although very few studies in the meta-analysis included multiple channels, it was possible to observe differences in the recognition accuracy of particular emotions through nonverbal channels across studies. It appears that some emotions are particularly well understood through particular nonverbal channels. Despite the fact that happiness was the most accurately understood emotion in the face, it was the least accurately understood emotion in the voice. Anger was the most accurately understood emotion in the voice, whereas it was relatively less well understood in the face. This suggests that different nonverbal channels do not merely carry redundant information but rather each may have certain specialized functions in the communication of emotion. Again, this finding reinforces the need for richer multichannel examinations of emotional expression and recognition.

### *Methodological Issues and Implications*

The results of the meta-analysis suggest the importance of a number of methodological issues and implications in cross-cultural research on emotion.

*Imitation versus posing.* Many studies used emotional stimuli created by imitating poses, photographs, or moving particular muscles, and the in-group advantage was statistically significant in such studies. However, when members of one cultural group were asked to imitate emotional expressions that originated from a cultural group other than their own, we often found null results in our test of in-group advantage (see Table 1 listings for Biehl et al., 1997; Boucher & Carlson, 1980; Kilbride & Yarczower, 1983; Matsumoto & Assar, 1992; Matsumoto & Ekman, 1988). The meta-analysis included three sets of studies in which non-Americans were asked to imitate facial expressions that had orig-

inated in the United States. In these three sets of studies, non-Americans were asked to pose their faces to be consistent with the FACS system of facial muscle movement developed by Ekman and colleagues (Ekman, Friesen, & Tomkins, 1971). This can be seen as asking the encoders to pose "American" faces. Although Ekman and colleagues have argued that the FACS system represents universal emotional expressions (e.g., Ekman, 1972), the system was created and validated using the facial expressions of American posers (Ekman et al., 1971). Rather than finding the typical in-group advantage, in these studies we often found an out-group advantage. In each case, Americans judged the foreign groups more accurately than their own members did. The results of these studies, however, would fit into the larger pattern if we reconsidered the coding of *in-group*. That is, if the in-group is the group whose expressions were imitated, then the results are consistent with other results in this meta-analysis. Another form of imitation was found in Yik et al. (1998), in which emotional constructs were imported from one culture to another, leading to anomalous results.

*Limitations.* The present findings are limited by the studies available for inclusion in this meta-analytic review. The research teams represented in the current studies have made many methodological choices criticized by other researchers and theorists. However, no particular criticized feature was present in all the studies that support the current results.

The collection of studies available for review was diverse in many ways—representing 42 nations, 23 ethnic groups, and 2 regional groups—but less diverse in other ways. Most important, the pool of samples tilted heavily toward stimuli originating in the United States and, to a lesser extent, Europe. This limitation made it crucial to consider the results of fully balanced studies separately, and to use the larger pool of studies only after confirming that their results were consistent with their balanced counterparts. Otherwise, we would risk confounding the presence of an in-group advantage with cultural differences in emotional expressive ability or emotion recognition ability. Another source of homogeneity is that most studies used multiple-choice response formats. Furthermore, most researchers sought to balance participants and stimuli expressors for gender, thus, the resulting restricted range made it difficult to examine gender differences in our analysis.

We regret that we were not able to examine other potentially fascinating moderators examined in recent studies on the cross-cultural appraisal of emotions, such as geopolitical divisions, climatic conditions, and economic factors (Scherer, 1997). We also regret not being able to analyze confusion analyses in order to examine the patterns of errors in the cross-cultural recognition of emotions because of the very small number of studies that report these matrices. Analyses of error patterns might be profoundly revealing of underlying mechanisms in emotional judgments. Such confusion matrices would also have allowed us to correct the present data for possible response bias that can alter the results of multiple-choice judgment studies (Wagner, 1993).

*Practical implications.* These findings have several practical implications. Many researchers make use of standardized stimuli that are purchased or borrowed from abroad. It will be valuable for individual researchers to consider whether such a practice is desirable given the goals of the particular study. One can infer from our results that using foreign stimuli would make less difference when studying emotions such as happiness or anger or dynamic

nonverbal channels of communication, which are associated with less of an in-group advantage. We hope that future cross-cultural studies of emotion strive to maximize the ecological validity of both the stimuli and judgment context.

### *Theoretical Implications: Emotional Dialects*

These findings have theoretical implications for the debate about universals and cultural differences in emotion. Emotions do not lose all meaning across cultural boundaries—but they may lose some meaning. These results echo work on linguistic and semantic structures that has found evidence for both universality and cultural specificity in semantic domains and cognitive representations. For example, Japanese and American individuals have been found to share semantic domains and cognitive representations of linguistic terms but also exhibit cultural specificity<sup>5</sup> (Romney, Boyd, Moore, Batchelder, & Brazill, 1996; Romney, Moore, & Rusch, 1997). In the search for explanations for cultural differences such as the in-group advantage, the present results emphasize the importance of developing a theory that addresses the interaction between the cultural background of the expresser and the perceiver of the emotion. Prior theories concerning individual characteristics of the culture of the expresser or the culture of the perceiver are less consistent with our findings that the match between expresser and perceiver accounts for cultural variability in emotion recognition.

Linguistic dialects refer to subordinate varieties of a basic language that vary in pronunciation, grammar, syntax, or vocabulary and are influenced by geographic, national, cultural, and social boundaries such as class and status (Francis, 1992; Romaine, 1994). Similar to linguistic dialects, the basic human language of emotional expression may have dialects that differ in the style of expression and interpretation. This meta-analysis suggests that emotional dialects might also be shaped by factors such as geographic, national, and social boundaries similar to linguistic dialects. In the same way that linguistic dialects influence how the same basic language can be shaped, expressed, and understood differently, we speculate that emotional dialects subtly nuance the manner in which emotions are universally expressed and understood.

<sup>5</sup> We thank David Matsumoto for alerting us to this work.

### References

- An asterisk indicates that the reference is a study included in the meta-analysis. A double asterisk indicates that the reference is a study or article used in the meta-analysis exclusively for normative data. A triple asterisk indicates a longer report containing additional data or description of methods for a published study included in the meta-analysis.
- \*Albas, D. C., McCluskey, K. W., & Albas, C. A. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology*, 7, 481–490.
- Allport, G. W., & Vernon, P. E. (1933). *Studies in expressive movement*. New York: Macmillan.
- Ambady, N., Bernieri, F., & Richeson, J. A. (2000). Towards a histology of social behavior: Judgmental accuracy from thin slices of behavior. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, 32, 201–272.

- Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology, 77*, 538–547.
- Anthony, T., Copper, C., & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality and Social Psychology Bulletin, 18*, 296–301.
- \*Bailey, W., Nowicki, S., Jr., & Cole S. P. (1998). The ability to decode nonverbal information in African American, African and Afro-Caribbean, and European American adults. *Journal of Black Psychology, 24*, 418–443.
- Baron, R. M., & Boudreau, L. (1987). An ecological perspective on integrating personality and social psychology. *Journal of Personality and Social Psychology, 53*, 1222–1228.
- \*Beier, E. G., & Zutra, A. J. (1972). Identification of vocal communication of emotions across cultures. *Journal of Consulting and Clinical Psychology, 39*, 166.
- \*Bhardwaj, A. (1982). *Nonverbal behavior: Face recognition and social distance of bilinguals and unilinguals*. Unpublished master's thesis, Ontario Institute for Studies in Education, Toronto, Canada.
- \*Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., et al. (1997). Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior, 21*, 3–21.
- \*Boggi-Cavallo, P. (1983). Il riconoscimento delle espressioni facciali delle emozioni: Uno studio cross-culturale [The recognition of facial expressions of emotion: A cross-cultural study]. In G. Atili & P. E. Ricci Bitti (Eds.), *Comunicare senza parole [Communicating without words]* (pp. 177–186). Rome: Bulzoni.
- \*Bond, C. F., Omar, A., Mahmoud, A., & Bonser, R. N. (1990). Lie detection across cultures. *Journal of Nonverbal Behavior, 14*, 189–204.
- \*Bormann-Kischkel, C., Hildebrand-Pascher, S., & Stegbauer, G. (1990). The development of emotional concepts: A replication with a German sample. *International Journal of Behavioural Development, 13*, 355–372.
- \*Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology, 11*, 263–280.
- Bruner, J. S., & Tagiuri, R. (1954). The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2, pp. 634–654). Cambridge, MA: Addison-Wesley.
- \*Buchman, J. S. (1973). Nonverbal communication of emotions among New York City cultural groups. (Doctoral dissertation, Columbia University, 1972). *Dissertation Abstracts International, 34*, 406–407.
- \*\*Bullock, M., & Russell, J. A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioral Development, 7*, 193–214.
- \*Caterina, R., Garotti, P. L., & Ricci Bitti, P. E. (1999). L'espressione non verbale del disprezzo [The nonverbal expression of contempt]. *Psychology and Society Review of Social Psychology, 25*, 11–23.
- \*Chan, D. W. (1985). Perception and judgment of facial expressions among the Chinese. *International Journal of Psychology, 20*, 681–692.
- \*Collins, M. H. (1996). Personality and achievement correlates of nonverbal processing ability in African-American children. (Doctoral dissertation, Emory University, 1996). *Dissertation Abstracts International, 57*, 1481.
- Deaux, K., & LaFrance, M. (1998). Gender. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 788–827). Boston, MA: McGraw-Hill.
- \*Demertzis, A., & Nowicki, S., Jr. (1996, April). *The ability of African-American and European-American college students to read emotions in same and other race facial expressions*. Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, GA.
- \*Dennis, D. G. (1982). Ethnicity and the capacity to interpret nonverbal communication. (Doctoral dissertation, Yeshiva University, 1982). *Dissertation Abstracts International, 43*, 1301.
- \*Dickey, E. C., & Knower, F. H. (1941). A note on some ethnological differences in recognition of simulated expressions of the emotions. *American Journal of Sociology, 47*, 190–193.
- \*Ducci, L., Arcuri, L., W/Georgis, T., & Sineshaw, T. (1982). Emotion recognition in Ethiopia: The effect of familiarity with Western culture on accuracy of recognition. *Journal of Cross-Cultural Psychology, 13*, 340–351.
- Efron, D. (1941). *Gesture and environment*. New York: King's Crown Press.
- \*Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (Vol. 19, pp. 207–282). Lincoln: University of Nebraska Press.
- Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115*, 268–287.
- \*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. (1976). *Pictures of facial affect* [Slides]. San Francisco: Department of Psychology, San Francisco State University.
- \*Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.
- Ekman, P., Friesen, W. V., & Tomkins, S. S. (1971). Facial Affect Scoring Technique (FAST): A first validity study. *Semiotica, 3*, 37–58.
- \*Ekman, P., Heider, E., Friesen, W. V., & Heider, K. (1972). *Facial expression in a preliterate culture*. Unpublished manuscript.
- \*Ekman, P., & Heider, K. G. (1988) The universality of a contempt expression: A replication. *Motivation and Emotion, 12*, 303–308.
- Ekman, P., & Rosenberg, E. L. (Eds.). (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. New York: Oxford University Press.
- \*Ekman, P., Sorensen, E. R., & Friesen, W. V. (1969, April 4). Pan-cultural elements in facial displays of emotions. *Science, 164*, 86–88.
- \*Evans, B. J., Coman, G. J., & Stanley, R. O. (1988). Scores on the Profile of Nonverbal Sensitivity—A sample of Australian medical students. *Psychological Reports, 62*, 903–906.
- \*Fernandez-Dols, J., 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.
- Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 915–981). Boston: McGraw-Hill.
- Francis, W. N. (1992). Dialectology. In W. Bright (Ed.), *International encyclopedia of linguistics* (pp. 349–355). Oxford, England: Oxford University Press.
- \*Gaebel, W., & Wolwer, W. (1992). Facial expression and emotional face recognition in schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience, 242*, 46–52.
- Galati, D., Scherer, K. R., & Ricci-Bitti, P. E. (1997). Voluntary facial expressions of emotion: Comparing congenitally blind with normal sighted encoders. *Journal of Personality and Social Psychology, 73*, 1363–1379.
- \*Gallois, C., & Callan, V. J. (1986). Decoding emotional messages: Influence of ethnicity, sex, message type, and channel. *Journal of Personality and Social Psychology, 51*, 755–762.
- \*Gates, G. (1923). An experimental study in the growth of facial perception. *Journal of Educational Psychology, 14*, 449–461.
- \*\*\*Gitter, A. G., & Black, H. (1968). *Perception of emotion: Differences*

- in race and sex of perceiver and expressor. (Rep. No. 17). Boston: Boston University, Communications Research Center.
- \*Gitter, A. G., Black, H., & Mostofsky, D. I. (1972a). Race and sex in the communication of emotion. *Journal of Social Psychology*, 88, 273–276.
- \*Gitter, A. G., Black, H., & Mostofsky, D. I. (1972b). Race and sex in the perception of emotion. *Journal of Social Issues*, 28, 63–78.
- \*Gitter, A. G., Kozel, N. J., & Mostofsky, D. I. (1972). Perception of emotion: The role of race, sex and presentation mode. *Journal of Social Psychology*, 88, 213–222.
- \*Gitter, A. G., Mostofsky, D. I., & Quincy, A. J., Jr. (1971). Race and sex differences in the child's perception of emotion. *Child Development*, 42, 2071–2075.
- Glass, G. V. (1978). Reply to Mansfield and Busse. *Educational Researcher*, 7, 3.
- \*Guidetti, M. (1991). Vocal expression of emotions: A cross-cultural and developmental approach. *Annee Psychologique*, 91, 383–396.
- \*Habel, U., Gur, R. C., Mandal, M. K., Salloum J. B., Gur, R. E., & Schneider, F. (2000). Emotional processing in schizophrenia across cultures: Standardized measures of discrimination and experience. *Schizophrenia Research*, 42, 57–66.
- \*Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition & Emotion*, 13, 225–266.
- \*Hall, C. W., Chia, R., & Wang, D. F. (1996). Nonverbal communication among American and Chinese students. *Psychological Reports*, 79, 419–428.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845–857.
- Hall, J. A. (1984). *Nonverbal sex differences: Accuracy of communication and expressive style*. Baltimore: Johns Hopkins University Press.
- Hall, J. A. (1998). How big are nonverbal sex differences? The case of smiling and sensitivity to nonverbal cues. In D. J. Canary & K. Dindia (Eds.), *Sex differences and similarities in communication: Critical essays and empirical investigations of sex and gender in interaction* (pp. 155–177). Mahwah, NJ: Erlbaum.
- Hall, J. A., & Friedman, G. B. (1999). Status, gender, and nonverbal behavior: A study of structured interactions between employees of a company. *Personality and Social Psychology Bulletin*, 25, 1082–1091.
- Hall, J. A., Halberstadt, A. G., & O'Brien, C. E. (1997). "Subordination" and nonverbal sensitivity: A study and synthesis of findings based on trait measures. *Sex Roles*, 37, 295–317.
- Harré, R. M. (Ed.). (1986). *The social construction of emotions*. Oxford, England: Basil Blackwell.
- \*Hatta, T., & Nachshon, I. (1988). Ear differences in evaluating emotional overtones of unfamiliar speech by Japanese and Israelis. *International Journal of Psychology*, 23, 293–302.
- Hecht, M. A., & LaFrance, M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin*, 24, 1332–1342.
- Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, 7, 119–137.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Henley, N. M. (1977). *Body politics: Power, sex, and nonverbal communication*. Englewood Cliffs, NJ: Prentice-Hall.
- \*Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, 288–299.
- 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, 1–22.
- \*Kellogg, W. N., & Eagleson, B. M. (1931). The growth of social perception in different racial groups. *Journal of Educational Psychology*, 22, 367–375.
- \*\*\*Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition & Emotion*, 10, 155–171.
- \*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.
- \*Kirouac, G., & Doré, F. Y. (1982). Identification des expressions faciales émotionnelles par un échantillon québécois francophone [Identification of emotional facial expressions in a French-speaking sample of Quebecois subjects]. *International Journal of Psychology*, 17, 1–7.
- \*Kirouac, G., & Doré, F. Y. (1983). Accuracy and latency of judgment of facial expressions of emotions. *Perceptual and Motor Skills*, 57, 683–686.
- \*Kirouac, G., & Doré, F. Y. (1985). Accuracy of the judgments of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior*, 9, 3–7.
- \*Kramer, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology*, 68, 390–396.
- \*Kretsch, R. A. (1968). *Communication of emotional meaning across national groups*. Unpublished doctoral dissertation, Columbia University, New York.
- \*Leung, J. P., & Singh, N. N. (1998). Recognition of facial expressions of emotion by Chinese adults with mental retardation. *Behavior Modification*, 22, 205–216.
- \*Machida, S. K. (1986). Teacher accuracy in decoding nonverbal indicants of comprehension and noncomprehension in Anglo- and Mexican-American children. *Journal of Educational Psychology*, 78, 454–464.
- \*Mandal, M. K., & Bhattacharya, B. B. (1985). Recognition of facial affect in depression. *Perceptual and Motor Skills*, 61, 13–14.
- \*Mandal, M. K., & Palchoudhury, S. (1989). Identifying the components of facial emotion and schizophrenia. *Psychopathology*, 22, 295–300.
- \*Mandal, M. K., Saha, G. B., & Palchoudhury, S. (1986). A cross-cultural study on facial affect. *Journal of Psychological Researches*, 30, 140–143.
- \*Markham, R., & Wang, L. (1996). Recognition of emotion by Chinese and Australian children. *Journal of Cross-Cultural Psychology*, 27, 616–643.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253.
- Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, 20, 92–105.
- 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. (1993). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an American sample. *Motivation and Emotion*, 17, 107–123.
- \*Matsumoto, D., & Assar, M. (1992). The effects of language on judgments of universal facial expressions of emotion. *Journal of Nonverbal Behavior*, 16, 85–99.
- \*Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE)*. [Slides]. San Francisco: Intercultural and Emotion Research Laboratory, Department of Psychology, San Francisco State University.
- \*Mazurski, E. J., & Bond, N. W. (1993). A new series of slides depicting facial expressions of affect: A comparison with the Pictures of Facial Affect Series. *Australian Journal of Psychology*, 45, 41–47.

- \*McAlpine, C., Kendall, K. A., & Singh, N. N. (1991). Recognition of facial expressions of emotion by persons with mental retardation. *American Journal on Mental Retardation*, *96*, 29–36.
- \*McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expression of emotion. *Journal of Cross-Cultural Psychology*, *17*, 211–224.
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238.
- \*McCluskey, K. W. (1980). Vocal communication of emotion: A program of research. (Doctoral dissertation, University of Manitoba, Canada, 1980). *Dissertation Abstracts International*, *41*, 599.
- \*McCluskey, K. W., & Albas, D. C. (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents and adults. *International Journal of Psychology*, *16*, 119–132.
- \*McCluskey, K., Albas, D., Niemi, R., Cuevas, C., & Ferrer, C. (1975). Cross-cultural differences in the perception of the emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Developmental Psychology*, *11*, 551–555.
- \*Mehta, S. D., Ward, C., & Strongman, K. (1992). Cross-cultural recognition of posed facial expressions of emotion. *New Zealand Journal of Psychology*, *21*, 74–77.
- Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112*, 197–204.
- Mesquita, B., Frijda, N. H., & Scherer, K. R. (1997). Culture and emotion. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology: Vol. 2. Basic processes and human development* (pp. 255–297). Boston: Allyn & Bacon.
- \*Niit, T., & Valsiner, J. (1977). Recognition of facial expressions: An experimental investigation of Ekman's model. *Acta et Commentationes Universitatis Tartuensis*, *429*, 85–107.
- \*Nowicki, S., Jr. (1997). *Manual for the receptive tests of the diagnostic analysis of nonverbal accuracy 2 (DANVA2)*. Atlanta, GA: Department of Psychology, Emory University.
- \*Nowicki, S., Jr., Glanville, D., & Demertzis, A. (1998). A test of the ability to recognize emotion in the facial expression of African American adults. *Journal of Black Psychology*, *24*, 333–348.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*, 208–224.
- O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An 'other-race effect' for categorizing faces by sex. *Perception*, *25*, 669–676.
- \*Ricci Bitti, P. E., Brighetti, G., Garotti, P. L., & Boggi-Cavallo, P. (1989). Is contempt expressed by pan-cultural facial movements? In J. P. Forgas & J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 329–339). Amsterdam: Elsevier.
- \*Ricci Bitti, P. E., Giovannini, D., Argyle, M., & Graham, J. (1979). La comunicazione di due dimensioni delle emozioni attraverso indici facciale e corporei [The communication of two dimensions of emotion expressed through the face and body]. *Giornale Italiano di Psicologia*, *6*, 341–350.
- \*Ricci Bitti, P. E., Giovannini, D., Argyle, M., & Graham, J. (1980). La comunicazione delle emozioni attraverso indici facciali e corporei [The communication of emotion expressed through the face and body]. *Giornale Italiano di Psicologia*, *7*, 85–94.
- Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury Press.
- Romaine, S. (1994). Dialect and dialectology. In R. E. Asher & J. M. Y. Simpson (Eds.), *The encyclopedia of language and linguistics* (pp. 900–907). Oxford, England: Pergamon Press.
- Romney, A. K., Boyd, J. P., Moore, C. C., Batchelder, W. H., & Brazill, T. J. (1996). Culture as shared cognitive representations. *Proceedings of the National Academy of Sciences*, *93*, 4699–4705.
- Romney, A. K., Moore, C. C., & Rusch, C. D. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *Proceedings of the National Academy of Sciences*, *93*, 5489–5494.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192.
- \*Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS Test*. Baltimore: Johns Hopkins University Press.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.
- Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, *92*, 500–504.
- Russell, J. A. (1991). Negative results on a reported facial expression of contempt. *Motivation and Emotion*, *15*, 285–292.
- Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*, 41–51.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*, 102–141.
- \*Russell, J. A., Suzuki, N., & Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expressions. *Motivation and Emotion*, *17*, 337–351.
- Russell, J. A., & Yik, M. S. M. (1996). Emotion among the Chinese. In M. H. Bond (Ed.), *The handbook of Chinese psychology* (pp. 166–188). Hong Kong: Oxford University Press.
- Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, *73*, 902–922.
- \*Scherer, K. R., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*, 76–92.
- Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, *66*, 310–328.
- \*\*\*Schneider, F., Gur, R. C., Gur, R. E., & Muenz, L. R. (1994). Standardized mood induction with happy and sad facial expressions. *Psychiatry Research*, *51*, 19–31.
- Seaford, H. W., Jr. (1975). Facial expression dialect: An example. In A. Kendon, R. M. Harris, & M. R. Key (Eds.), *Organization of behavior in face-to-face interaction*. The Hague, the Netherlands: Mouton.
- \*Shimoda, K., Argyle, M., & Ricci Bitti, P. E. (1978). The intercultural recognition of emotional expressions by three national social groups: English, Italian and Japanese. *European Journal of Social Psychology*, *8*, 169–179.
- \*Sogon, S., & Masutani, M. (1989). Identification of emotion from body movements: A cross-cultural study of Americans and Japanese. *Psychological Reports*, *65*, 35–46.
- \*Sorensen, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (Ed.), *Psychological anthropology* (pp. 361–372). Chicago: Aldine.
- \*Sprouse, A., Boyd, B., Hall, C., Webster, R., & Bolen, L. (1995, April). *Age, race and gender based nonverbal social perception*. Paper presented at the meeting of the Southeastern Psychological Association, Savannah, GA.
- \*Stokes, D. R. (1984). Nonverbal communication: Race, gender, social class, world view and the PONS test: Implications for the therapeutic dyad. (Doctoral dissertation, Ohio State University, 1983). *Dissertation Abstracts International*, *44*, 3544.
- \*Streit, M., Wolwer, W., & Gaebel, W. (1997). Facial-affect recognition and visual scanning behavior in the course of schizophrenia. *Schizophrenia Research*, *24*, 311–317.
- \*Strong, K. (1978). Children's ability to recognize facial expressions of



- affect. (Doctoral dissertation, Memphis State University, 1978). *Dissertation Abstracts International*, 39, 3564–3565.
- \*Sweeney, M. A., Cottle, W. C., & Kobayashi, M. J. (1980). Nonverbal communication: A cross-cultural comparison of American and Japanese counseling students. *Journal of Counseling Psychology*, 27, 150–156.
- TeleGeography*. (1997). Washington, DC: Author.
- \*Toner, H. L., & Gates, G. R. (1985). Emotional traits and recognition of facial expression of emotion. *Journal of Nonverbal Behavior*, 9, 48–66.
- \*Van Bezooijen, R., Otto, S., & Heenan, T. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology*, 14, 387–406.
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, 17, 3–28.
- \*Wallbott, H. G. (1991a). Recognition of emotion from facial expression via imitation—Some indirect evidence for an old theory. *British Journal of Social Psychology*, 30, 207–219.
- \*Wallbott, H. G. (1991b). The robustness of communication of emotion via facial expressions: Emotion recognition from photographs with deteriorated pictorial quality. *European Journal of Social Psychology*, 21, 89–98.
- Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, 78, 105–119.
- \*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. (1980). *A multicultural project: The development of an interracial facial recognition test Phase III, a comparison between New Canadian West Indians' and Canadians' sensitivity to interracial facial expressions and social distance*. Unpublished manuscript, Ontario Institute for Studies in Education, Toronto, Canada.
- \*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.
- \*Wolwer, W., Streit, M., Polzer, U., & Gaebel, W. (1996). Facial affect recognition in the course of schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, 246, 165–170.
- \*Yik, M. S. M., Meng, Z., & Russell, J. A. (1998). Adults' freely produced emotion labels for babies' spontaneous facial expressions. *Cognition & Emotion*, 12, 723–730.
- \*Yik, M. S. M., & Russell, J. A. (1999). Interpretation of faces: A cross-cultural study of a prediction from Fridlund's theory. *Cognition & Emotion*, 13, 93–104.

## Appendix

### Sample Calculations

#### Example 1: Ekman and Heider (1988)

This study is an example in which members of a single cultural in-group, the Minangkabau of West Sumatra, judged emotional expressions produced both by members of their own group as well as expressions produced by members of out-groups from the United States and Japan.

Judgment accuracy data for out-group Americans and Japanese is reported in Table 1 of Ekman and Heider (1988). The average accuracy of the photos listed in the chart is 83.5%. Because respondents could choose from a list of seven emotions (contempt, disgust, anger, fear, sadness, surprise, and happiness), we apply the error correction formula with a base rate due to guessing of 1/7. Subtracting 1/7 (14.3%) from 83.5%, and then dividing it by the new possible total of 6/7 (85.7%), we get a corrected value of 80.8%. Because this is the accuracy of the cross-cultural judgment, 80.8% is the universal accuracy.

Judgment accuracy for in-group Minangkabau photographs is reported in Table 2 of Ekman and Heider (1988). The average accuracy of the two photographs is 92.0%. The same guessing correction gives us an in-group accuracy of 90.7%. The in-group advantage is the difference between in-group and out-group accuracy, 90.7% – 80.8%, which is 9.9%.

#### Example 2: Hatta and Nachshon (1988)

This study is an example in which the emotional expressions of a single group, the Japanese, are judged both by in-group Japanese participants and are also judged by out-group Israeli participants.

Out-group Israeli subjects obtained accuracy of 90.1% and 90.3% in their left and right ears, respectively. Because there were two choices (negative and positive), we correct these scores with a guessing base rate of .5 and then average them together for an overall out-group accuracy score of 80.4%. The in-group Japanese scores of 93.1% and 91.2% for the left and right ears, respectively, are subjected to the same guessing correction and averaged to get an in-group accuracy score of 84.3%. The in-group advantage is the in-group Japanese score minus the out-group Israeli score, 84.3% – 80.4%, which is 3.9%.

Because information is given on the standard deviations of these scores, a significance test can also be performed on the in-group advantage. First, we take the average across the original accuracy data for the left and right ear for each group to get 92.2% for the Japanese and 90.2% for the Israelis. We pool the standard deviations ( $s$ ) across the ears using the formula,  $s^2_{\text{pooled}} = ((s_1^2(n_1 + 1) + s_2^2(n_2 - 1)) / ((n_1 - 1) + (n_2 - 1)))$ , (Rosenthal & Rosnow, 1991), to get 10.1% for the Japanese and 10.7% for the Israelis. We then pool the standard deviations for the Japanese and Israelis together using the same formula and take the square root to get an overall 10.1%. We then calculated  $t$  as,  $t = ((x_1 - x_2) / (S_{\text{pooled}} * \sqrt{(n_1 * n_2) / (n_1 + n_2)}))$ , (Rosenthal & Rosnow, 1991), to get  $t(69) = 0.81, p = .21$ .

Received November 15, 1999

Revision received August 15, 2001

Accepted August 21, 2001 ■