Emotion
2002, Vol. 2, No. 1, 75–84

1528-3542/02/$5.00 DOI: 10.1037//1528-3542.2.1.75

BRIEF REPORT

Cross-Cultural Patterns in Emotion Recognition: Highlighting Design and Analytical Techniques

Hillary Anger Elfenbein
Harvard University

Manas K. Mandal
Indian Institute of Technology, Kharagpur

Nalini Ambady
Harvard University

Susumu Harizuka
Kyushu University

Surender Kumar
Chikushi Women University

This article highlights a range of design and analytical tools for studying the cross-cultural communication of emotion using forced-choice experimental designs. American, Indian, and Japanese participants judged facial expressions from all 3 cultures. A factorial experimental design is used, balanced $n \times n$ across cultures, to separate "absolute" cultural differences from "relational" effects characterizing the relationship between the emotion expressor and perceiver. Use of a response bias correction is illustrated for the tendency to endorse particular multiple-choice categories more often than others. Treating response bias also as an opportunity to gain insight into attributional style, the authors examined similarities and differences in response patterns across cultural groups. Finally, the authors examined patterns in the errors or confusions that participants make during emotion recognition and documented strong similarity across cultures.

The long-standing debate between psychologists who argue whether emotions are universal versus culturally specific is ceding to attempts to integrate evidence for both perspectives (e.g., Fiske, Kitayama, Markus, & Nisbett, 1998; Markus & Kitayama, 1991; Mesquita & Frijda, 1992; Mesquita, Frijda, & Scherer, 1997; Scherer & Wallbott, 1994). Researchers studying emotion recognition have extensively documented both cultural similarities and differences. Cross-cultural studies of emotion recognition have been one of several central sources of evidence in favor of emotional universality. Classic studies (e.g., Ekman, 1972; Ekman, Sorenson, & Friesen, 1969; Izard, 1971) demonstrated that facial photographs of Americans expressing basic emotions could be recognized at above-chance accuracy in literate and preliterate cultures. Despite this evidence for universality, the same studies also provide evidence for cultural differences, given that American samples generally outperformed others when viewing these American stimuli. For example, in Izard's (1971) large-scale study, American and European groups correctly identified 75%–83% of the facial photographs, whereas the Japanese group scored 65% and the African group correctly identified only 50%.

Hillary Anger Elfenbein, Program in Organizational Behavior, Harvard University; Manas K. Mandal, Department of Humanities and Social Sciences, Indian Institute of Technology, Kharagpur, Kharagpur, India; Nalini Ambady, Department of Psychology, Harvard University; Susumu Harizuka, Faculty of Education, Center for Clinical Psychology and Human Development, Kyushu University, Fukuoka City, Japan; Surender Kumar, Department of Preschool Education, Chikushi Women University, Fukuoka City, Japan.

Preparation of this article was supported by a National Science Foundation Graduate Student Fellowship, a Rand research grant, Presidential Early Career Award for Scientists and Engineers Grant BCS-9733706, and Indian Institute of Technology Scheme for Innovative Research and Development Grant IIT-Accts. 05-19-05. We thank Susan Choi, Heather Gray, Abby Marsh, and Kevyn Yong for their helpful comments, and Sanaz Ghazal, Maiga Miranda, and Julie Weintraub for research assistance.

Correspondence concerning this article should be addressed to Hillary Anger Elfenbein, Harvard Business School, Baker Library 477, Soldiers Field, Boston, Massachusetts 02163. E-mail: hillary@post.harvard.edu

"Forced-Choice" Experimental Designs

The majority of studies on the cross-cultural communication of emotion have used "forced-choice" designs, in which participants view expressions and enter multiple-choice responses from a predetermined list. The relative merits of such studies have been hotly debated (e.g., Ekman, 1994; Frank & Stennett, 2001; Haidt & Keltner, 1999; Russell, 1994). Russell (1993, 1994) argued that these studies prime participants to interpret stimuli as expressions of emotion, portray choices as mutually exclusive categories, and inflate agreement across cultures by constraining choices. He demonstrated that forced choice can show consensus on clearly incorrect categories when relevant choices are missing from the list. Participants may choose a particular label when prompted by a multiple-choice list, even if they would not have generated the same label under free-response conditions (Rosenberg & Ekman, 1995; Wagner, 2000). Furthermore, response choices are often generated by North American researchers, whose lists may or may not correspond to emotion constructs in other cultures (Haidt & Keltner, 1999; Russell & Yik, 1996).

However, there are also reasons for researchers to continue using forced-choice designs. First, there are theoretical considerations. Forced-choice designs fit theory that emotions are categorical in nature because individual emotions may have evolved separately to solve distinct problems for human survival (Ekman, 1992; Izard, 1994). By contrast, Russell (1980) maintained that emotions are qualitatively similar to each other, yet vary quantitatively along underlying dimensions such as valence and arousal. Forced-choice designs also facilitate the testing of recognition rather than production of emotion-related constructs, which are different abilities evolving separately (Izard, 1994). Theoretical clarity may also decrease the vulnerability of forced-choice studies to methodological artifact, as Izard (1994) argued that theory-based lists of responses are less likely to miss desired choices and artificially boost consensus on an incorrect choice. A second consideration in favor of forced-choice designs is empirical, as findings using forced-choice designs generally replicate those using alternative methods such as free-response and dimensional ratings (e.g., Boucher & Carlson, 1980; Ekman, 1994; Haidt & Keltner, 1999; Izard, 1994). Haidt and Keltner (1999) and Frank and Stennett (2001) replicated high agreement levels with modified forced-choice designs featuring improvements such as an expanded list of emotion words as well as a "none-of-the-above" option. Third, perhaps the likeliest reason forced-choice designs remain is convenience. Dimensional ratings can be tedious for participants, and analyzing open-ended participant responses can be time-consuming and ambiguous for experimenters.

Analytical Possibilities With Forced-Choice Experiments

Given the popularity of forced-choice designs, it is helpful to consider analytical tools that offer richer perspectives on the data that are collected. Often, researchers using forced-choice studies restrict themselves to examining whether recognition accuracy is greater than that predicted by chance guessing alone, or differences in accuracy across samples. These restrictions have notable exceptions, and the goal of the current article is to highlight a range of analytical tools researchers have used to gain greater insight from forced-choice designs.

Interactions, in Addition to Main Effects

Broadly speaking, two different types of theories can explain cross-cultural differences in emotional communication. The first type is *absolutist,* examining the fixed attributes of the groups expressing emotion or of the groups perceiving emotion. For example, Matsumoto (1989, 1992) and Schimmack (1996) have argued that collectivistic cultural groups are less accurate at expressing and perceiving negative emotion, because of the risk of disrupting social order. These arguments predict absolute differences in communication accuracy across groups, or "main effects" in terms of an analysis of variance (ANOVA). By contrast, *relational* theories characterize the relationship between the cultures of the emotional expressor and perceiver rather than the fixed attributes of one group or the other. They focus on match or similarity rather than on absolute characteristics of cultures. In terms of an ANOVA, relational theories predict an interaction between expressor and perceiver cultural groups. Recent empirical work has provided initial support for relational perspectives on emotional communication (Elfenbein & Ambady, 2001), including a meta-analysis of the cross-cultural literature on emotion recognition suggesting the presence of an in-group advantage (Elfenbein & Ambady, 2002a, 2002b; but see also Matsumoto, 2002). That is, emotional communication may be more accurate when the expressor and the perceiver are both members of the same cultural group rather than members of different cultural groups.

Both absolutist and relational perspectives can contribute toward a valuable understanding of emotional communication. However, only factorial experimental designs balanced $n \times n$ across cultural groups are capable of separating absolutist from relational effects, to examine each of these independent terms while controlling for the other. In balanced studies, members of each culture view stimuli equally from members of their own and members of other groups in the study. Thus, they allow the calculation of absolutist main effects for expressors and for perceivers, as well as the interaction term representing relational effects. The first purpose of this study is to examine evidence for a relational effect by using a balanced design, predicting a replication of the in-group advantage:

> Hypothesis 1: There is an in-group advantage in emotional communication.

Correcting for Response Bias

In studying judgments of nonverbal behavior, response scoring has been an important measurement issue (Rosenthal, 1987; Scherer, Banse, & Wallbott, 2001; Wagner, 1993). Researchers conducting judgment studies generally analyze percentage of accuracy "hit rates" (Wagner, 1993), which are diagonal entries in a confusion matrix plotting correct answer categories in one dimension and participant judgments in the other. Hit rates indicate the proportion of occasions participants labeled stimuli with the intended category. An extreme example can illustrate why using hit rates exclusively can bias the results of judgment studies. Sorenson (1975) reported that members of the Bahinemo tribe of New Guinea reacted to Ekman, Sorenson, and Friesen's (1969) photographs of facial expressions by labeling all of the stimuli as "angry." Although participants did correctly identify anger expressions 100% of the time, they could not distinguish it from other emotions. To correct for such variability in the use of response categories, Wagner (1993) proposed an "unbiased hit rate," the hit rate multiplied by one minus the rate of false alarms, then normalized with an arcsin transformation. This correction is similar to signal detection methods except, unlike signal detection terms, it allows separate analyses for each stimulus category (Wagner, 1993).

Meta-analyses and other reviews reflect the methodologies of the studies they include. Most studies in Elfenbein and Ambady's (2002b) recent meta-analysis of emotion recognition did not provide the confusion matrices necessary to calculate unbiased hit rates. However, the majority of these studies fit Wagner's (1993) criteria for relatively low impact from response bias: using posed expressions and stimuli preselected for high accuracy rates. The current study illustrates the use of unbiased hit rates by replicating the in-group advantage while correcting for response bias, with the following prediction:

> Hypothesis 2: There is an in-group advantage in emotion using unbiased hit rates.

Learning From Response Bias

Instead of treating response bias exclusively as a problem, researchers can also examine variability in the use of response categories as an opportunity to learn about attributional style in emotional judgments. Although the traditional focus has been on sensitivity in nonverbal communication, judgment biases and errors can be as revealing as accuracy (DePaulo & Friedman, 1998). Tomkins and McCarter (1964) noted individual differences in participant tendencies to endorse certain emotions. They argued that frequent personal experience with particular emotions can prime individuals toward those emotional states, and they are more likely to attribute those emotions to stimuli whether or not the particular state is present. Indeed, research has validated the relationship between emotional response bias and important social phenomena, notably between hostile attribution bias and children's social adjustment. Aggressive boys have a general bias toward seeing anger and threat in the nonverbal behavior of others (Crick & Dodge, 1994; Nasby, Hayden, & DePaulo, 1980; Schultz, Izard, & Ackerman, 2000). Moreover, Barth and Bastiani (1997) demonstrated that children's emotion recognition biases were actually more relevant than their recognition accuracy in predicting social behavior.

The current study explores whether patterns of response bias in emotion recognition may differ across cultures. Previous research suggests specific predictions. Matsumoto (1989, 1992) argued that collectivistic cultural groups avoid recognizing negative emotion in order to preserve social order. By contrast, individualistic societies like the United States tolerate and even may encourage the perception of negative emotion. This suggests that collectivist groups would demonstrate patterns of response bias in which they are less likely to attribute negative emotions to others. Furthermore, this tendency would be especially strong when judging members of their own cultural group, with whom social order would be more important. This study examines differences in response bias tendencies and the positive versus negative valence of response bias in particular, with the following predictions:

> Hypothesis 3: Valence response bias in emotion recognition will be more positive among perceivers from collectivistic rather than individualistic cultures.
>
> Hypothesis 4: Valence response bias will be more positive when judging emotions expressed by members of the perceiver's own cultural group.

Examining Confusion Profiles

Researchers have also examined systematic patterns in misattributing emotions. Certain emotions may be mistaken at random, whereas others may be mistaken only for specific other emotions. These patterns provide insight into judgments by suggesting the cues used to infer emotions (Banse & Scherer, 1996). Tomkins and McCarter (1964) noted that their participants showed systematic similarities in error patterns, tending to confuse particular emotions with each other, for example, surprise for fear and disgust for anger. Some confusions never occurred, for example, shame for happiness and vice versa. They argued that emotions are confused with others sharing similar expressive qualities. More recently, Haidt and Keltner (1999) examined misattributions across basic emotions, additional emotions such as embarrassment and amusement, and other expressions such as covering one's face. Arguing that universality is characterized by the discreteness of the modal response choice, in addition to cross-culturally high recognition accuracy, they used error profiles as one source of evidence that expressions ranged in their degree of universality. Scherer et al. (2001) demonstrated strong cross-cultural agreement in systematic error patterns for vocal expressions, albeit with slightly lower agreement levels for the country most culturally disparate from the others they sampled. They argued that it would be worthwhile to replicate this work by using facial expressions, as well as by using a balanced design with several cultural groups, which is the final purpose of this study:

> Hypothesis 5: Patterns of emotion recognition errors are substantially similar across cultures.

Thus, the overall goal of this article is to illustrate a variety of analytical techniques for use with forced-choice judgment studies. We do so by using a new data set, balanced with Indian, Japanese, and American participants, each judging stimuli from all three groups. Because various collectivistic cultures can differ in their emotional tendencies (e.g., Camras et al., 1998), we include members from two different collectivistic societies.

Method

Emotional Expressions

Stimuli were black-and-white photographs of facial expressions. Researchers developed the photographs while residing in the nation from which the posers originated. From each cultural group, we used one man and one woman posing each of the following: anger, disgust, fear, happiness, neutral, sadness, and surprise. To match practice effects for individual posers, the set from each culture contained an average of 7 individual people posing in two photographs each. Posers from each group were similar in age.

India. Mandal (1987) instructed posers to imagine an emotional situation and to pose an appropriate expression for each emotion. Two separate consensus samples of Indian judges with at least 70% agreement validated recognition levels. Intensity ratings were at least three points higher for the intended emotion than any other emotion, using a 7-point scale.

Japan. Mandal, Harizuka, Bhushan, and Mishra (2001) developed these stimuli by using the same technique and validation methods as the Indian set, with Japanese participants.

United States. Ekman and Friesen's (1976) Pictures of Facial Affect collection was created to conform with their theoretical model for the appearance of prototypical facial expressions of emotion, using the Facial Affect Coding System (FACS; Ekman & Friesen, 1978). Participants moved their facial muscles according to specific instructions rather than attempting to pose particular emotions. In spite of this methodological difference, we used these photographs because of their wide popularity in emotion recognition research. Given that Ekman and Friesen (1978) developed their model for prototypical expressions within the United States, it is arguably consistent with American norms for appropriate facial expressions. A consensus sample of Americans judged each photograph used in the current study, with the exception of neutral, with accuracy of 88% or greater.

Participants

Perceivers from each country represented in the stimulus materials were university students (India, *n*

= 48: 30 males and 18 females; Japan, $n = 80$: 29 males and 51 females; United States, $n = 68$: 28 males and 40 females). American participants had non-Asian ancestry (African American, $n = 5$, 7%; Hispanic, $n = 12$, 18%; Caucasian, non-Hispanic, $n = 51$, 75%).

Judgment Tasks and Procedure

Participants viewed the 42 photographs of facial expressions on a computerized task programmed with SuperLab (1997). Photographs appeared on the screen at 5 in. (12.7 cm) wide and 7.5 in. (19 cm) tall, with a resolution of 72 pixels per inch (28.3 pixels per cm), in a randomized order differing for each participant. Participants first judged two photographs in familiarization trials not used in analyses. Each photograph remained on the screen until the participant entered a permitted response. Response choices used the language of instruction of the students' university, English in India and the United States and Japanese in Japan (translations: *ikari* [anger], *kyoufu* [fear], *kenno* [disgust], *yorkobi* [happiness], *chuurich* [neutral], *kanashimi* [sadness], and *odoroki* [surprise]). The experimenter in each nation was a member of the same cultural group as the participants.

Results

Cultural Differences in Emotion Recognition Accuracy

Unbiased hit rates were calculated with Wagner's (1993) formula and the confusion matrix[1] for each individual participant and analyzed with a 3 (expresser culture) × 3 (perceiver culture) × 2 (perceiver sex) × 7 (emotion) ANOVA. Accuracy varied across judge cultures, $F(2, 190) = 16.2$, $p < .01$. Tukey's post hoc tests revealed that American participants scored higher than Indian participants, who scored higher than Japanese participants. Consistent with past findings on gender differences (Hall, 1978), female participants were more accurate than male participants, $F(1, 190) = 5.0$, $p < .03$, $r = .16$, although there were no interactions between other factors and judges' gender. Accuracy varied across expressor cultures, $F(2, 380) = 283.8$, $p < .01$. Tukey's post hoc tests revealed that American expressors were more easily understood than Indian expressors, who were more easily understood than Japanese expressors. There was also an effect across emotions, $F(6, 1140) = 256.3$, $p < .01$, such that happiness and neutrality were judged the most accurately, and fear and anger the least. Table 1 lists mean levels of unbiased accuracy across cultures and, for comparison, Table S4, (see Footnote 1) lists the same using conventional hit rates.

Table 1
Unbiased Hit Rate (H_u) Emotion Recognition Accuracy Mean Values and Interaction Across Nations of Expressors and Judges

	Expressor			
Perceiver	India	Japan	USA	Total
Mean values				
India	.738	.378	.885	.667
Japan	.556	.397	.730	.561
USA	.716	.450	1.009	.725
Total	.670	.408	.874	.651
Interaction effect				
India	**.052**	(.046)	(.006)	
Japan	(.024)	**.079**	(.055)	
USA	(.028)	(.032)	**.060**	

Note. Bold typeface indicates in-group advantage for judgments of members of the same cultural group. Results in parentheses indicate negative values.

In support of Hypotheses 1 and 2, which predict an in-group advantage in using conventional and unbiased hit rates, respectively, there was an interaction between judge and expressor cultures for conventional hit rates, $F(4, 380) = 11.7$, $p < .01$, and unbiased hit rates, $F(4, 380) = 10.6$, $p < .01$. The in-group advantage uses this interaction term along with contrast weights of $\lambda = +2$ for in-group judgments and $\lambda = -1$ for out-group judgments, as per the prediction that residuals show higher accuracy for judging members of one's own cultural group. The correlation between these contrast weights and the interaction residuals, "$r_{alerting}$" (Rosenthal & Rosnow, 1991) provides an indication of the fit between residuals and the predicted contrast. The $r_{alerting}$ of .91 using conventional hit rates and .95 using unbiased hit rates suggests that this fit is strong. In support of Hypotheses 1 and 2, this contrast representing the in-group advantage was significant for conventional hit rates, $F(1, 380) = 42.7$, $p < .01$, $r_{contrast} = .32$, and unbi-

[1] All supplementary tables referenced in this article, including Tables S1, S2, and S3 containing confusion matrices for Indian, Japanese, and American emotional expressions, respectively, are available via the Internet at www.people.hbs.edu/helfenbein. Note that we list effect sizes for all significance tests aside from omnibus F tests consisting of unfocused comparisons among conditions. Readers without access to the Internet can contact Hillary Anger Elfenbein for a copy of the supplementary tables.

ased hit rates, $F(1, 380) = 40.4$, $p < .01$, $r_{\text{contrast}} = .31$, with effect sizes considered medium in magnitude (Rosenthal & Rosnow, 1991).

Cultural Differences in Response Bias

We used a 3 (expressor culture) × 3 (perceiver culture) × 2 (perceiver sex) × 7 (emotion) ANOVA to analyze patterns in response category usage. Note that all main effects for between-subjects terms are zero, as the forced-choice design of the study required that usage across categories total exactly 100% for each participant. Table S5 (see Footnote 1) lists response usage across emotions, judge culture, judge gender, and expressor culture. Participant responses varied across emotions, $F(6, 1140) = 86.2$, $p < .01$. Judges more often endorsed expressions as neutral and happy, and less often as afraid or sad. There was an interaction between emotion and judge culture, $F(12, 1140) = 4.1$, $p < .01$. Japanese participants relatively less often endorsed expressions as afraid and more often as surprised, compared with participants from other groups. Americans relatively more often endorsed expressions as afraid and less often as surprised or disgusted. Unlike the above analysis of accuracy, in response bias there was an interaction between perceiver gender and emotion, $F(6, 1140) = 2.3$, $p < .04$, such that men relatively more often endorsed photographs as neutral and less often as sad. There was also an interaction between emotion and expressor culture, $F(12, 3086) = 289.3$, $p < .01$. Japanese expressions were relatively more often judged as happy and less often as afraid, and American expressions were more often judged as angry and less often as neutral.

Valence response bias is each participant's tendency to endorse emotional expressions as more positive versus negative than the intended category. We coded anger, disgust, fear, and sadness as negative, neutrality and surprise as neutral, and happiness as positive. Each participant's score for valence response bias was the sum of:

> 2 multiplied by the proportion of negative expressions rated as positive, 1 multiplied by the proportion of negative expressions rated as neutral, 1 multiplied by the proportion of neutral expressions rated as positive, −1 multiplied by the proportion of neutral expressions rated as negative, −1 multiplied by the proportion of positive expressions rated as neutral, and −2 multiplied by the proportion of positive expressions rated as negative.

Thus, participants always rating emotions within their intended valence receive a score of zero, those rating all emotions as positive receive a score of +3, and those rating all emotions as negative receive a score of −3. This score accounts for the uneven distribution of stimuli across the three valences, although the uneven distribution does add to measurement error. Table 2 summarizes values across cultural groups, and Tables S6, S7, and S8 (see Footnote 1) list greater detail. Overall, valence response bias was weakly positive ($M = 0.27$, scale from −3 to +3).

Table 2
Valence Response Bias Mean Values and Interaction Across Nations of Expressors and Judges

Perceiver	Expressor: India	Japan	USA	Total
Mean values				
India	.12	.74	(.01)	.28
Japan	.16	.61	.11	.29
USA	.07	.72	(.05)	.25
Total	.12	.69	.02	.27
Interaction effect				
India	**(.01)**	.05	(.04)	
Japan	.02	**(.09)**	.06	
USA	(.02)	.07	**(.04)**	

Note. Bold typeface indicates valence response bias for judgments of members of the same cultural groups. Results in parentheses indicate negative values. Positive values indicate that emotional expressions were endorsed with categories more positive than intended, and negative values indicate endorsement with categories more negative than intended.

In contrast with Hypothesis 3, there was no effect for perceiver culture, $F(2, 190) = 0.1$, *ns*. That is, cultural groups did not differ in bias toward endorsing emotions as more positive than intended. There was a large effect for expressor culture, $F(2, 380) = 223.2$, $p < .01$, such that Japanese photographs were generally rated as more positive than Indian or American. There was also an interaction between perceiver and expressor cultures, $F(4, 380) = 3.6$, $p < .01$. The correlation between these interaction residuals and the contrast weights distinguishing between in-group ($\lambda = +2$) and out-group ($\lambda = -1$) judgments was $r_{\text{alerting}} = -.64$. In contrast with Hypothesis 4, participants were actually more positive in their response bias with members of other cultural groups. This contrast was significant, $F(1, 380) = 12.4$, $p < .01$, $r_{\text{contrast}} = .18$, albeit with a small effect size (Rosenthal & Rosnow, 1991). This effect appears to result exclusively from the Japanese stimuli and participants. Analyzing Indian and American data separately, there is no interaction between perceiver and expressor cultures, $F(1, 112) = 0.0$, *ns*.

Confusion Patterns Across Cultures

Table 3 lists correlations among error profiles. These error profiles consist of the confusion matrices from Tables S1, S2, and S3 (see Footnote1) with the diagonal entries removed. In support of Hypothesis 5, there is substantial agreement across cultures. The average correlation was .84 for error profiles containing a common group of expressors and was .60 for profiles based on different expressor groups. This suggests that different cues contained within stimuli may be partially responsible for variations in error patterns.

In addition to error patterns, it is also worthwhile to examine agreement in accuracy patterns. An accuracy profile is the vector of accuracy values across all the individual emotions tested. Tables S9 and S10 (see Footnote 1) list correlations among accuracy profiles using unbiased and conventional hit rates, respectively. Although conventional hit rates display strong agreement, unbiased hit rates show great variability. Unbiased hit rates had average correlation of .66 for profiles containing a common group of perceivers, and −.16 for profiles based on different perceiver groups. This suggests that some perceiver groups are systematically more accurate with some emotions than others. This notion is supported by a significant interaction between perceiver culture and emotion in the earlier analysis of unbiased hit rates, $F(12, 1140) = 18.3$, $p < .01$. The difference between Tables S9 and S10 (see Footnote 1) suggests that response bias may hide cultural differences when analyzing conventional hit rates, and may overstate universality.

Discussion

These results highlight the value for researchers studying emotion recognition with forced-choice designs to use a range of additional methods and analytical techniques beyond simple percentage of accuracy. First, looking beyond absolute effects that can characterize cultural differences in emotional communication, we examined a relational effect using the interaction between the cultural groups of expressors and perceivers. Second, we used error patterns to apply Wagner's (1993) response bias correction to accuracy levels. Using corrected values, we replicated the in-group advantage (Elfenbein & Ambady, 2002b), which suggests that the effect does not result from artifactual bias in the use of response categories. This replication does not imply that response bias is unimportant, but rather that it is not responsible for previous findings of this particular effect using this particular methodology. Posed expressions preselected for high accuracy tend to be the least vulnerable to response bias problems (Wagner, 1993). However, as we discuss below, in the current study one result did differ markedly across corrected versus uncorrected values. Calculating unbiased hit rates is tedious but by no means difficult. It is worthwhile for researchers to use corrected values and to reanalyze data sets to confirm that prior findings do not result from response artifacts.

Third, we treated response bias as an opportunity to learn more about attributional style. We found evidence for cross-cultural similarities in the willingness to endorse particular emotions, as well as some dif-

Table 3
Correlations of the Error Profiles Across Combinations of Expressor and Perceiver Nations

	1.	2.	3.	4.	5.	6.	7.	8.
Indian expressors								
1. Indian judges								
2. Japanese judges	**.86**							
3. United States judges	**.93**	**.85**						
Japanese expressors								
4. Indian judges	<u>.75</u>	.69	.60					
5. Japanese judges	.79	<u>.78</u>	.61	**.94**				
6. United States judges	.78	.71	<u>.69</u>	**.94**	**.88**			
United States expressors								
7. Indian judges	<u>.58</u>	.55	.54	<u>.48</u>	.54	.49		
8. Japanese judges	.65	<u>.87</u>	.65	.52	<u>.64</u>	.49	**.59**	
9. United States judges	.40	.53	<u>.48</u>	.30	.33	<u>.29</u>	**.54**	**.66**

Note. Bold typeface indicates correlations between confusion matrices sharing the same nation of expressors, and underlined typeface indicates correlations between confusion matrices sharing the same nation of judges.

ferences. We investigated more closely a specific prediction, but found no cross-cultural differences in perceivers' tendency to endorse emotions as more positive than they actually are. In contrast with prediction, participants had greater positive bias when judging members of out-group cultures. We speculate they may have been polite in attributing greater positivity to expressions they found ambiguous or confusing.

Fourth, the current study provides additional evidence for cultural universals and differences in patterns of error and accuracy. As Scherer et al. (2001) found with vocal expression, there was substantial agreement in the patterns of emotion recognition errors when judging facial expressions. This was particularly true for profiles involving a common set of expressors. Making common errors with common stimuli suggests that perceivers use many pan-cultural cues and decision rules. However, accuracy patterns across emotions—at least using unbiased hit rates—appeared to differ across cultural groups of perceivers. Because uncorrected hit-rate accuracy values displayed broad universality in these patterns, the response bias correction may be worthwhile to avoid possibly overstating the degree of universality in patterns of emotion recognition accuracy.

This study has important limitations that future research should address. First and most important, the methods of creating stimuli varied. Although all matched across age, gender, and practice effects, the Indians and Japanese attempted to pose emotions, whereas the Americans posed specific muscle movements. This likely accounts for higher recognition rates of American photographs. The participant groups also differed, as American perceivers attended a nationally ranked liberal arts school, Indian perceivers attended a nationally ranked science school, and Japanese perceivers attended a school not nationally ranked. This may help account for cultural main effects and in some cases very low recognition rates. This study also examined emotion recognition outside of its social and functional context. The asocial nature of viewing facial expressions on a computer screen may understate the role of norms arising in social context (Buck, 1984). Additionally, this study used static, posed black-and-white photographs of facial expressions. Stimuli capturing more closely the dynamic, full-color, full-channel nature of emotional expressions would bolster ecological validity. Another limitation is that cultural group membership was apparent to participants through the facial features of the emotional stimuli. This confounds possible motivational differences and stereotypes about members of different cultures (Hess, Senecal, & Kirouac, 1996), with effects possibly due to subtle differences in emotional expressive style.

Future research should further theory that can account for empirical observations of differences in emotional communication across cultures. For example, Elfenbein and Ambady (2002a, 2002b) used a speculative metaphor of emotional "dialects"—whereby emotional expression is a universal language, with dialects that may differ across cultures—that awaits direct testing. Stimulus materials should be unconstrained to allow cultural differences to emerge, and these spontaneous or matched culturally consistent emotional expressions could be judged with a balanced design.[2] Analyzing judgments in terms of a lens model (Brunswick, 1955; Gifford, 1994) could reveal how specific differences in emotional expression map onto differences in perception.[3]

The current article emphasizes support for emerging theories of emotional communication that integrate evidence for both universality and cultural specificity. In empirically evaluating such theories, it is helpful for researchers to use a range of analytical techniques that weave together diverse strands of evidence.

---

[2] We thank Klaus Scherer for suggesting this design for a future study.

[3] We thank an anonymous reviewer for this suggestion.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70,* 614–636.

Barth, J. M., & Bastiani, A. (1997). A longitudinal study of emotion recognition and preschool children's social behavior. *Merrill-Palmer Quarterly, 43,* 107–128.

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

Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217.

Buck, R. (1984). *The communication of emotion.* New York: Guilford Press.

Camras, L. A., Oster, H., Campos, J., Campos, R., Ujiie, T., Miyake, K., Wang, L., & Meng, Z. (1998). Production of emotional facial expressions in European American, Japanese, and Chinese infants. *Developmental Psychology, 34,* 616–628.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101.

DePaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In. D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology: Vol. 2.* (4th ed., pp. 3–40). Boston: McGraw-Hill.

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

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. (1976). *Pictures of facial affect* [Slides]. San Francisco: Department of Psychology, San Francisco State University.

Ekman, P., & Friesen, W. V. (1978). *Facial Affect Coding System: A technique for the measurement of facial movement.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969, April 4). Pancultural elements in facial displays of emotions. *Science, 164,* 86–88.

Elfenbein, H. A., & Ambady, N. A. (2001). *Cultural similarity's consequences: A relational perspective on cross-cultural differences in emotion recognition.* Manuscript submitted for publication.

Elfenbein, H. A., & Ambady, N. (2002a). Is there an in-group advantage in emotion recognition? *Psychological Bulletin, 128,* 243–249.

Elfenbein, H. A., & Ambady, N. (2002b). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 205–235.

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.

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology, 80,* 75–85.

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology, 66,* 398–412.

Haidt, J., & Keltner, D. (1999). Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. *Cognition and Emotion, 13,* 225–266.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85,* 845–857.

Hess, U., Senecal, S., & Kirouac, G. (1996). Recognizing emotional facial expressions: Does perceived sociolinguistic group make a difference? *International Journal of Psychology, 31,* 93.

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.

Mandal, M. K. (1987). Decoding of facial emotions, in terms of expressiveness, by schizophrenics and depressives. *Psychiatry, 50,* 371–376.

Mandal, M. K., Harizuka, S., Bhushan, B., & Mishra, R. C. (2001). Cultural variation in hemifacial asymmetry of emotion expressions. *British Journal of Social Psychology, 40,* 385–398.

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

Matsumoto, D. (2002). Methodological requirements to test a possible in-group advantage in judging emotions across cultures: Comment on Elfenbein and Ambady (2002) and Evidence. *Psychological Bulletin, 128,* 236–242.

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.* (pp. 255–297). Needham Heights, MA: Allyn & Bacon.

Nasby, W., Hayden, B., & DePaulo, B. M. (1980). Attributional bias among aggressive boys to interpret unambiguous social stimuli as hostile displays. *Journal of Abnormal Psychology, 89,* 459–468.

Rosenberg, E., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion, 19,* 111–138.

Rosenthal, R. (1987). *Judgment studies: Design, analysis, and meta-analysis.* Cambridge, England: Cambridge University Press.

Rosenthal, R., & Rosnow, R. L. (1991). Essentials of behavioral research: Methods and data analysis (2nd ed.). New York: McGraw-Hill.

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

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., & Yik, M. S. M. (1996). Emotion among the Chinese. In M. H. Bond (Ed.), *The handbook of Chinese psychology* (pp. 166–188). Hong Kong, China: Oxford University Press.

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.

Schimmack, U. (1996). Cultural influences on the recognition of emotion by facial expressions: Individualistic or Caucasian cultures? *Journal of Cross-Cultural Psychology, 27,* 37–50.

Schultz, D., Izard, C. E., & Ackerman, B. P. (2000). Children's anger attribution bias: Relations to family environment and social adjustment. *Social Development, 9,* 284–301.

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

SuperLab Pro 1.04 [Computer software]. (1997). Phoenix, AZ: Cedrus Corporation.

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

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3–28.

Wagner, H. L. (2000). The accessibility of the term "contempt" and the meaning of the unilateral lip curl. *Cognition and Emotion, 14,* 689–710.

Received August 29, 2001
Revision received January 1, 2002
Accepted January 1, 2002 ■