

# Emotion Recognition Across Cultures: The Influence of Ethnicity on Empathic Accuracy and Physiological Linkage

José Angel Soto  
Pennsylvania State University

Robert W. Levenson  
University of California, Berkeley

The present study tested whether empathic accuracy and physiological linkage during an emotion recognition task are facilitated by a cultural match between rater and target (cultural advantage model) or unaffected (cultural equivalence model). Participants were 161 college students of African American, Chinese American, European American, or Mexican American ethnicity. To assess empathic accuracy—knowing what another person is feeling—participant's (raters) used a rating dial to provide continuous, real-time ratings of the valence and intensity of emotions being experienced by 4 strangers (targets). Targets were African American, Chinese American, European American, or Mexican American women who had been videotaped having a conversation with their dating partner in a previous study and had rated their own feelings during the interaction. Empathic accuracy was defined as the similarity between ratings of the videotaped interactions obtained from raters and targets. To assess emotional empathy—feeling what another person is feeling—we examined physiological linkage (similarity between raters' and targets' physiology). Our findings for empathic accuracy supported the cultural equivalence model, while those for physiological linkage provided some support for the cultural advantage model.

*Keywords:* empathic accuracy, empathy, cultural advantage, physiological linkage

Emotions play an essential role in human communication. An important part of our interpersonal lives is the production, perception, interpretation, and response to emotional signals. Being able to perceive these signals accurately carries clear advantages for predicting behavior as well as forming and maintaining social bonds. These proximal functions impact more distal goals such as survival and reproduction in humans and in other species (Preston & de Waal, 2002). Thus, compelling advantages accrue to those individuals who can accurately read the emotions of others. These advantages may be even greater in situations when the individuals are members of the same in-group (Anderson & Keltner, 2002), suggesting a possible cultural advantage in emotion detection.

The question of whether individuals are more accurate at recognizing the emotional displays of members of their own culture versus those of other cultures has been an important issue in emotion research for several decades. Findings of same-culture facilitation and same-culture equivalence have both been reported (e.g., Ekman et al., 1987; Elfenbein & Ambady, 2002a; Matsu-moto, 2002). One characteristic of this work has been its reliance on static emotional stimuli (e.g., still photographs, vignettes, etc.) and one-time emotional judgments (e.g., a photograph of a face that is identified as “angry”). Such stimuli are common in the

laboratory; however, real-world emotion judgments are typically made more continuously as emotions unfold dynamically over time. In the present study, we revisit the question of a cultural advantage using a methodology that allows us to assess dynamic emotion judgments that arguably have greater ecological validity.

## Cultural Equivalence Versus Cultural Advantage Models

### *Cultural Equivalence Model*

The cultural equivalence model of emotion recognition asserts that the mechanisms underlying the expressive and receptive aspects of emotional communication are deeply rooted in evolution (Darwin, 1872, 1965). It has now been over a century since Darwin (1872, 1965) noted striking similarities across species (including humans) in how critical emotional information such as anger/aggression, happiness, and fear is communicated in postural and expressive signals. These similarities in conveying emotional information across species are consistent with an evolutionary argument that within many social animals there appears to be a natural ability to relay emotional signals efficiently and effectively with one another—a position that has been subsequently endorsed by others (Chevalier-Skolnikoff, 1974; Linnankoski, Laasko, Aulanko, & Leinonen, 1994). By extension, our ability to recognize the emotional signals of all conspecifics (not just a subset of these) carries great value for the ultimate survival of the species (Preston & de Waal, 2002). Thus, a cultural equivalence model predicts that individuals should be equally accurate in understanding the emotions of in-group and out-group members.

### *Empirical Evidence*

Empirical support for the cultural equivalence model can be found in a number of studies. For example, Ekman and colleagues

---

José A. Soto, Department of Psychology, Pennsylvania State University; Robert W. Levenson, Department of Psychology, University of California, Berkeley.

This research was supported by National Institute on Aging Grant R37-AG17766 and National Institute on Mental Health Grant MH39895 to Robert W. Levenson.

Correspondence concerning this article should be addressed to José A. Soto, Department of Psychology, Pennsylvania State University, University Park, PA 16802. E-mail: josésoto@psu.edu

(Ekman, 1972; Ekman & Friesen, 1971; Ekman et al., 1987; Matsumoto, 1993) demonstrated that people from different cultures are able to identify correctly the emotions portrayed in photographs of emotional facial expressions. This evidence extends to members of preliterate, visually isolated tribes in New Guinea and Borneo who were able to match emotion labels to pictures of facial expressions posed by Caucasians (Ekman, Sorenson, & Friesen, 1969). Although such data are widely interpreted as evidence for the universality of emotional facial expressions, this conclusion is not without controversy (see Ekman, 1994; Izard, 1994; Russell, 1994, 1995, for a lively debate regarding this conclusion). Nevertheless, the above studies and many others demonstrate that cultural differences between raters (i.e., observers or judges who are making inferences about another's emotions) and targets (those whose emotions are being rated or judged) do not hinder accurate recognition of emotions (Boucher & Carlson, 1980; Gitter, Kozel, & Mostofsky, 1972; Kilbride & Yarczower, 1983).

### *Cultural Advantage Model*

Anderson and Keltner (2002) pointed out that evolutionary forces may also favor a cultural advantage in recognizing emotions. In support for this notion, O'Toole, Peterson and Deffenbacher (1996) found that individuals process the visual characteristics of same-race faces more accurately and efficiently than other-race faces. Moreover, the mere perception that one belongs to the same group as a target can lead to increased in-group accuracy in emotion recognition (Thibault, Bourgeois, & Hess, 2006). Belonging to the same cultural group may also increase accuracy by virtue of familiarity with slight differences in the ways emotions are expressed by members of that group, an idea captured by the notion of nonverbal accents proposed by Marsh, Elfenbein, and Ambady (2003) and the dialect theory of communicating emotion (see Elfenbein, Beaupré, Lévesque, & Hess, 2007).

### *Empirical Evidence*

Several early studies documented increased accuracy in recognizing the emotions from others from the same culture (Freeman, 1984; Vinacke & Fong, 1955; Wolfgang & Cohen, 1988); however, these findings were often overshadowed by findings of "universality" within the same studies (e.g., Vinacke & Fong, 1955). Subsequent studies have increasingly emphasized the findings of a cultural/ethnic advantage in recognizing the emotions of others. For instance, Albas, McCluskey, and Albas (1976) compared the ability of White Canadians and Indian Canadians to judge the emotional tone of content-filtered speech from either White or Indian speakers. Although both groups were able to correctly judge the emotions of the other group with above-chance accuracy, each group was better at this task when rating speech from their own cultural group. Freeman (1984) found that preschool boys scored higher on a measure of cognitive empathy when judging targets of the same ethnicity than those of a different ethnicity. Wolfgang and Cohen (1988) found that Anglo Canadians were better able to recognize the facial expressions of Anglo Canadians than those of West Indian Canadians and vice versa. A similar pattern of results was found for Chinese and Australian participants by Markham and Wang (1996).

This in-group advantage has been found for ethnic minorities as well as majority group members (Weathers, Frank, & Spell, 2002). In a meta-analysis of 97 studies, Elfenbein and Ambady (2002b) found a significant in-group advantage such that emotion recognition accuracy was 9.3% better when participants judged expressions posed by members of the same cultural group. Although this conclusion has been contested (see Elfenbein & Ambady, 2002a; Matsumoto, 1993), there is certainly sufficient evidence of a cultural advantage for this to constitute a rival hypothesis to the cultural equivalence model.

## Mechanisms Underlying Empathic Accuracy: Cognitive and Emotional Empathy

Empathic accuracy is the ability of one person to recognize the thoughts and feelings of another person at a specific moment in time (Ickes, 1993). It is one of the important subareas subsumed under the larger rubric of "empathy research" (which also includes such topics as prosocial behavior, sympathy, and compassion). Empathic accuracy is a measure of what people can do in the realm of emotion recognition. It does not tell us how they do it. In terms of how we come to know what others are feeling a critical distinction has been made between cognitive empathy and emotional empathy (Davis, 1994).

### *Cognitive Empathy*

Cognitive empathy involves "knowing" what the target individual is feeling. At the core of cognitive empathy is some form of perspective taking or conscious cognitive processes (Hatfield, Cacioppo & Rapson, 1994) that allows an observer to deduce logically what a target is feeling. More important, observers do not have to become emotionally aroused for cognitive empathy to occur; they can simply process available cues and information and come to a conclusion as to what the other person is feeling. In the present study, our assessment of empathic accuracy is viewed as a measure of cognitive empathy. It determines how well individuals from various ethnic groups can tell how strangers from their own and other ethnic groups are feeling, but does not tell us how this occurs.

### *Emotional Empathy*

Emotional empathy involves "feeling" what the target individual is feeling. Applied to empathic accuracy, it suggests that one way that we come to know what another person is feeling is by experiencing some version of the emotion ourselves. One basis for emotional empathy is emotional contagion, a process whereby an observer comes to share the emotions of a target as a result of automatic mimicry and synchronization with the target's facial expressions, vocalizations, postures, and movements (see Dimberg, 1982; Eisenberg et al., 1989; Hatfield et al., 1994; Preston & de Waal, 2002). "Mirror neurons" provide another basis for emotional empathy. These neurons, respond to both the generation of certain behaviors as well the observation of those behaviors being performed by others (Rizzolatti & Craighero, 2004; Rizzolatti, Fadiga, Fogassi, & Gallese, 1999). Originally observed in nonhuman primates, there is now indirect evidence from functional imaging studies for their existence in humans (Decety & Jackson,

2004; Decety & Lamm, 2006; Decety & Meyer, 2008; but see counter evidence in Lingnau, Gesierich, & Caramazza, 2009).

In prior research, researchers have explored "physiological linkage," a state in which the patterns of autonomic and somatic physiological response of an observer mimic those of the person being observed. They initially demonstrated that this occurs both between spouses (Levenson & Gottman, 1983) and within spouses (Gottman & Levenson, 1985) when married couples viewed videotapes of their interactions and rate each other's emotions. It was also later demonstrated that physiological linkage occurs when individuals observe and rate the emotions of strangers (Levenson & Ruef, 1992), especially when the observers show emotional facial expressions (Soto, Pole, McCarter, & Levenson, 1998). We and others (Preston & de Waal, 2002) consider physiological linkage as a form of emotion contagion in which observers have emotional reactions that are similar to those of the target person and that these similar emotions produce similar patterns of autonomic and peripheral nervous system activity (e.g., Ekman, Levenson, & Friesen, 1983; Levenson, 2003).

### *Emotional Empathy and Empathic Accuracy*

Support for the notion that emotional empathy is related to empathic accuracy comes from several sources. It has been shown that greater physiological linkage was associated with greater accuracy for detecting targets' negative emotions (Levenson & Ruef, 1992). More recently, Zaki, Weber, Bolger, and Ochsner (2009) demonstrated that accuracy in rating emotions was correlated with activity in brain regions associated with shared sensorimotor representations (inferior parietal lobule and bilateral dorsal premotor cortex) as might be expected during emotional empathy and with activity in regions indicative of mental state attributions (medial prefrontal cortex and superior temporal sulcus) as might be expected during cognitive empathy.

### Hypotheses

The present study addresses the issue of whether there is a cultural advantage in two aspects of empathy. First, there is a test of cognitive empathy using an empathic accuracy paradigm in which raters rate targets' emotions continuously in real time during

a naturalistic social interaction. Second, there is a test of emotional empathy that assesses the extent of physiological linkage (similarity between raters' and targets' physiological responses) during this rating task. This design enables us to test rival predictions between the cultural equivalence and cultural advantage models. Specifically, the cultural equivalence model predicts no differences in empathic accuracy or physiological linkage as a function of the match or mismatch between raters' and targets' ethnicity. In contrast, the cultural advantage model predicts greater empathic accuracy and physiological linkage when raters' and targets' ethnicity is the same as compared to when they are different.

## Method

### *Participants (Raters)*

We recruited 161 full-time college students from the greater San Francisco Bay area colleges and universities to serve as emotion raters. These participants were of either African American ( $n = 39$ ), Chinese American ( $n = 43$ ), European American ( $n = 38$ ), or Mexican American ( $n = 41$ ) ethnicity. Within each of the four ethnic groups, approximately 60% of the participants were women and 40% were men. The mean age of the sample was 20.23 years ( $SD = 2.44$ ) and it did not differ significantly across the four ethnic groups. Sixteen individuals failed to report their age.

Participants were recruited by making announcements in classes, posting flyers, and inviting students in common areas to participate. All interested individuals were screened for eligibility via telephone, Internet, or in person. Specific eligibility criteria for each ethnic group, presented in Table 1, were established in consultation with researchers who had special expertise relevant to these ethnic groups. The criteria aimed to ensure an equal amount of cultural exposure to American culture and to the participant's culture of origin (see Soto, Levenson, & Ebling, 2005, for a complete discussion of these criteria). Therefore, for the purposes of this study we use the terms *ethnicity* and *culture* interchangeably. Those who qualified for the study were scheduled for a laboratory session. Participants also completed a questionnaire package but these data were not used in the present study. Some students participated in the study in exchange for psychology

Table 1  
*Eligibility Criteria for Ethnic Groups*

Criterion variable	European American	Chinese American	African American	Mexican American
Parents and grandparents birthplace	United States	China, Taiwan, or Hong Kong	United States	Mexico
Religious affiliation	Christian or Catholic	None	None	Catholic
Language fluency	English	Mandarin or Cantonese and English	English	Spanish and English
Ethnic identification	None	None	Moderate identification with African American culture	None
Close friends		At least 50% of friends while growing up of same ethnicity		
Neighborhood		At least 10% of neighborhood while growing up of same ethnicity		

*Note.* All participants were born in the United States. Those criteria that span across columns indicate criteria applicable to all groups. Religious criteria for Mexican Americans and European Americans were considered met if participants indicated they were currently practicing the appropriate religion or grew up practicing the religion. Ethnic identification for African Americans was determined using five brief questions about their exposure and identification with African American culture.

course credit whereas others were paid \$35 for participating in the research.

### *Apparatus and Materials*

**Audiovisual.** A remote-controlled, high-resolution video camera, partially concealed behind darkened glass, was used to obtain frontal views of each participant's face and upper torso as they completed their ratings. All instructions were presented visually on a 13" TV monitor placed directly in front of the participant at a distance of 4 feet (1.22 m). Participants were informed that they would be videotaped.

**Physiology.** Seven physiological functions were measured using a system consisting of a Grass Model 7 12-channel polygraph (Grass Instruments, Quincy, MA) and a microcomputer: (a) Cardiac interbeat interval, was measured using Beckman miniature electrodes (Beckmann Instruments, Fullerton, CA) with Redux paste placed on opposite sides of the chest. The interval between successive R waves was measured in milliseconds. (b2) Pulse transmission time to the finger was measured using a UFI photoplethysmograph (UFI Instruments, Morro Bay, CA) attached to the top phalange of the second finger of the nondominant hand. Transmission time was measured as the time interval between the R wave of the electrocardiogram and the upstroke of the peripheral pulse at the finger. (c) Finger pulse amplitude, the trough-to-peak amplitude of the finger pulse, was also measured from the UFI photoplethysmograph. (d) Ear pulse transmission time was measured using a UFI photoplethysmograph attached to the right ear lobe. Transmission time was measured between the R wave of the electrocardiogram and the upstroke of the peripheral pulse at the ear. (e) Skin conductance level, the level of sweat-gland activity on the surface of the hand was measured by passing a constant voltage between Beckman regular size electrodes attached to the palmar surface of the lower phalanges of the first and second fingers on the nondominant hand. The electrodes were filled with an electrolyte paste consisting of sodium chloride in unibase. (f) Finger temperature was measured using a Yellow Springs Instruments thermistor (Yellow Springs, OH) attached to the palmar surface of the first phalange of the fourth finger of the nondominant hand. (g) General somatic activity was measured using an electro mechanical transducer attached to a platform under the participant's chair. The transducer generated an electrical signal, measured in arbitrarily designated units, proportional to the amount of movement in any direction. The seven measures that were examined are sensitive to cardiac, vascular, and electrodermal responses thought to be highly relevant to emotion.

**Rating dial.** The emotion rating dial consists of a mechanical dial that traverses a 180° arc over a 9-point scale with the labels *extremely negative* at one end, *neutral* in the middle, and *extremely positive* at the other end. The dial is attached to a potentiometer in a voltage-dividing circuit that provides a signal to the computer from which the precise dial position can be determined (Levenson & Ruef, 1997). This rating dial was used to obtain ratings from the participants in the current study and was used previously by target individuals to rate their own emotions as part of their original study protocol (see below).

**Stimulus materials (targets).** Videotapes of eight targets interacting with their dating partners were selected as stimulus materials for the current study. Two targets (and their partners)

represented each of the four ethnic groups represented by our participants (African American, Chinese American, European American, Mexican American) and met the same recruitment criteria presented in Table 1. These videotapes had been collected previously as part of a study of cultural influences on couples' interactions. In that study, couples came to the laboratory and engaged in three 15-min conversations about: (a) the events of the day, (b) a conflict area in their relationship, and (c) a pleasant area in their relationship. During the conversations, their physiological responses were measured (the same responses used in the present study and described above). Following each conversation, partners watched the videotape of their conversation and used the rating dial described above to provide continuous ratings of their own feelings during the conversation (these procedures were originally developed by Levenson & Gottman, 1983, for use with married couples). This previous study provided approximately 180 videotaped conversations from which we selected our final targets.

Several steps were taken to select the final eight target videotapes. First, we decided to use only female partners as targets given research indicating that women are more expressive than men (Hall, Carter, & Horgan, 2000; LaFrance & Banaji, 1992) and engender greater empathic accuracy than men (Klein & Hodges, 2001; Levenson & Ruef, 1992). Second, to ensure the videos had sufficient emotion, we made use of the self-ratings provided by the female partners in the original conversation (see data reduction below) to select only those in which the female partner rated her own feelings as negative or positive for more than half the time. Third, to ensure a balance between positive and negative affect, we selected only those conversations in which the female partner's ratings showed a ratio of negative to positive affect between 0.5 and 2.0. For conversations meeting these three criteria, we then had the female partner's emotions rated by a panel of six judges (using the same procedures described below) and excluded conversations that were deemed too easy or too difficult to rate in terms of the judges' average empathic accuracy scores.

Finally, from among the remaining videotaped conversations we selected two targets from each ethnic group; one in which the events of the day was the topic of conversation and one in which a conflict area or pleasant area of the relationship was the topic of conversation. When more than one conversation was available, we made the final selection at random. The final eight target videotapes were randomly divided into two sets of four targets (distributed equally across ethnic group and conversation type) and participants were randomly assigned to view one of these sets, with the order of the four targets within each set counterbalanced across participants.

### *Procedures*

Participants came to the laboratory for a 2½-hr experimental session. To minimize discomfort, the participants had contact only with a research assistant of their own gender while the physiological sensors were being attached. Participants were seated in a chair facing a monitor and written consent was obtained. Afterward, a research assistant attached the physiological sensors described earlier and explained their functions.

Participants were told that they would be watching four videotaped conversations between dating couples and that they would be

asked to use the rating dial to provide a continuous report of how they thought the woman in the video was feeling. To make sure they understood how to operate the dial, participants were asked to move it to indicate positive, negative, and neutral emotions. The videorecordings showed a split screen image with each partner occupying half of the screen. Before each conversation the appropriate side of screen was covered so that only the female partner's image was visible (participants could still hear the male partner). Following each conversation, participants were asked how likable and attractive they found the female target, and how engaged they were by the conversation.<sup>1</sup> These latter questions were answered using a 0 (*not at all*) to 8 (*extremely*) scale.

### Data Reduction

*Empathic accuracy.* Procedures for deriving empathic accuracy and physiological linkage scores are based on those employed by Levenson and Ruef (1992). The 10-s averages were standardized and each period was classified as positive, negative, or neutral using both raw and standardized scores. To be coded as a positive period, the raw rating dial score had to be greater than or equal to 6.0 (referenced to the original 1 to 9 emotion rating dial scale) and the  $z$  score had to be greater than or equal to 0.5. For a period to be coded negative, the raw rating dial score had to be less than or equal to 4.0 and the  $z$  score had to be less than or equal to  $-0.5$ .

Empathic accuracy scores were determined separately for positive and negative emotion using a sequential analysis (Levenson & Gottman, 1983) that compares the participant's ratings with that target's ratings for two lags: agreement between participant and target in the same 10-s period (lag zero) and agreement between target's rating in one 10-s period and participant's rating in the following 10-s period (lag one). This analysis produces a  $z$  score index of how likely it is that the participant's ratings agree with the target's ratings. Thus, each participant had four empathic accuracy scores for each target they rated: positive accuracy scores for lag zero and lag one and negative accuracy scores for lag zero and lag one.

*Linkage.* Using a computer program written by the second author, each of the physiological measures was quantified and averaged into successive 10-s periods. Physiological linkage was determined by applying a bivariate time-series analysis (Gottman, 1981) evaluating the extent of relatedness between the physiological data collected from the participant while rating the interaction and the physiological data collected from target while in the original interaction. Linkage was computed in both directions, in each case controlling for autocorrelations (i.e., cyclicity) within each time series (if uncontrolled, these autocorrelations can unduly inflate estimates of relationships across time series). The analysis yields log likelihood statistics, with approximate chi-square distributions for each physiological variable. Following procedures used previously with these kinds of data (Levenson & Gottman, 1983; Levenson & Ruef, 1992), this process was carried out for each of the seven physiological variables that were common to participants and targets in both directions of prediction (i.e., participant to target, target to participant). The proportion of these 14 statistics that was statistically significant was used as an index of overall physiological linkage between rater and target.

## Results

### Overall Empathic Accuracy

Before addressing the principal hypotheses, we computed a simple estimate of the overall level of rating accuracy of all our participants. This index, computed as the percentage of 10-s periods rated the same by the participant and target, was computed separately for positive periods and negative periods at lag zero and lag one and for each target ethnicity. The resulting 16 accuracy scores are presented in Table 2. The mean level of overall empathic accuracy, collapsing across participant ethnicity, ranged from 13% to 45% with only the overall accuracy ratings of African Americans exceeding chance (set at 33%, because a period could be positive, negative, or neutral). Accuracy levels for our sample were consistent with those reported by Levenson and Ruef (1992).

### Empathic Accuracy: Testing for Cultural Advantage

Participants' empathic accuracy scores were analyzed using a  $4 \times 4$  (Rater Ethnicity  $\times$  Target Ethnicity)<sup>2</sup> multivariate analysis of variance (MANOVA) with target ethnicity as a repeated measure. The dependent variables were participants' empathic accuracy scores for positive and negative emotion periods at lag zero and at lag one. If ethnic match between rater and target is associated with higher empathic accuracy, a significant interaction between rater ethnicity and target ethnicity in the overall MANOVA would be expected. Results revealed no significant interactions between Rater Ethnicity  $\times$  Target Ethnicity,  $F(36, 1744.3) = 0.84$ ,  $ns$ ,  $\eta_p^2 = .02$ . Thus, we found no support for an in-group advantage in empathic accuracy.

To explore further the findings reported above, follow-up univariate analyses of variances (ANOVAS) were conducted separately for each of the four dependent variables: positive accuracy scores at lags of zero and one and negative accuracy scores at lags of zero and one. Across each of the dependent variables, the effect of the Rater Ethnicity  $\times$  Target Ethnicity interaction was not significant, further bolstering the lack of an in-group advantage. These findings, however, only indicate the lack of an absolute cultural advantage. Matsumoto (2007) distinguished between an absolute cultural advantage and a relative culture advantage. The former is determined by a test of the omnibus interaction in a balanced ANOVA design (every rater culture rates every target culture) such as that utilized in this study.

Following recommendations outlined by Matsumoto (2007), we tested for a relative cultural advantage by first running a linear mixed models procedure that allows for the specification of a main effects-only model given a repeated-measures design. This analy-

<sup>1</sup> Of the 161 total participants, only 117 (73%) were asked about their level of engagement with the conversation as this question was introduced after the first 43 subjects had been run. The percentage of each participant ethnic group that did not answer this question is as follows: African Americans (10%), Chinese Americans (32%), European Americans (50%), and Mexican Americans (17%).

<sup>2</sup> Identical analyses with participant gender included as a between-subjects variable were conducted for each dependent variable and the pattern of findings was unchanged. Therefore, for the ease of presentation all analyses and findings presented will exclude gender as a variable of interest.

Table 2  
Overall Empathic Accuracy

Accuracy variable, target ethnicity	% of time target's affect rating was matched by participant's rating		
	<i>M</i>	<i>SE</i>	Range
Positive affect—EA			
Lag zero	20	2	0–92
Lag one	19	2	0–92
Negative affect—EA			
Lag zero	26	2	0–91
Lag one	26	2	0–91
Positive affect—CA			
Lag zero	28	2	0–100
Lag one	28	2	0–100
Negative affect—CA			
Lag zero	30	2	0–81
Lag one	32	2	0–85
Positive affect—AA			
Lag zero	37	2	0–90
Lag one	36	2	0–95
Negative affect—AA			
Lag zero	45	3	0–100
Lag one	43	3	0–100
Positive affect—MA			
Lag zero	13	1	0–78
Lag one	13	1	0–78
Negative affect—MA			
Lag zero	29	2	0–80
Lag one	31	2	0–78

Note. EA = European American; Lag zero = target and subject gave rating in the same 10-s period; lag one = target's rating in a given 10-s period was matched by the subject's rating in the following period; CA = Chinese American; AA = African American; MA = Mexican American.

sis was repeated for each of the four empathic accuracy variables and the resulting residuals were saved. For each dependent variable, we then ran a single-degree of freedom planned contrast that tested the hypothesis that among the matrix of 16 residuals representing the Rater Ethnicity  $\times$  Target Ethnicity interaction, participants rating targets of the same ethnicity (matrix diagonal) would show equally high empathic accuracy scores (four contrast weights of one) compared to those participants rating targets of a different ethnicity (12 contrast weights of  $-1/3$  along the off-diagonals). Across all four dependent variables, the planned contrast analysis was not significant, indicating the lack of a relative cultural advantage as well: positive accuracy at lag zero,  $t(497.2) = 0.62$ , *ns*; positive accuracy at lag one,  $t(499.9) = 0.63$ , *ns*; negative accuracy at lag zero,  $t(499.9) = -0.53$ , *ns*; and negative accuracy at lag one,  $t(501.3) = 0.07$ , *ns*.<sup>3</sup>

#### Empathic Accuracy: Main Effects of Rater and Target Ethnicity

We also were interested in the presence of any main effects associated with rater ethnicity and target ethnicity. With all four dependent variables (positive and negative accuracy at lags one and zero) in the model, main effects emerged for both rater ethnicity,  $F(12, 405.1)^4 = 1.88$ ,  $p < .05$ ,  $\eta_p^2 = .05$  and target ethnicity,  $F(12, 1230.6) = 24.2$ ,  $p < .001$ ,  $\eta_p^2 = .17$ . Follow-up univariate analyses failed to demonstrate a significant rater ethnic-

ity effect for any of the four accuracy variables. The effect of target ethnicity, however, was significant across each of the four accuracy scores: positive accuracy at lag zero,  $F(2.71, 423.4) = 31.04$ ,  $p < .001$ ,  $\eta_p^2 = .17$ ; positive accuracy at lag one,  $F(2.70, 421.5) = 31.19$ ,  $p < .001$ ,  $\eta_p^2 = .17$ ; negative accuracy at lag zero,  $F(2.66, 415.46) = 37.36$ ,  $p < .001$ ,  $\eta_p^2 = .19$ ; and negative accuracy at lag one,  $F(2.59, 403.2) = 30.57$ ,  $p < .001$ ,  $\eta_p^2 = .16$ . Table 3 lists the participants' mean accuracy scores, averaging across rater ethnicity, for each of the four target ethnicities. The pattern of means for each of the four empathic accuracy scores reflects a clear and consistent pattern. All participants were most accurate in rating the emotions of the African American targets, followed by Chinese Americans, European Americans, and Mexican Americans, in order from most accurately rated to least accurately rated.

#### Physiological Linkage

Mean physiological linkage scores (presented in Table 4) were analyzed using a  $4 \times 4$  (Rater Ethnicity  $\times$  Target Ethnicity) repeated measures ANOVA, with target ethnicity serving as a repeated measure and physiological linkage (percentage of the 16 linkage variables that was significant) as the independent variable. If ethnic match between rater and target leads to greater physiological linkage, a significant interaction between rater ethnicity and target ethnicity in the overall ANOVA would be expected. Results showed no main effect for rater ethnicity,  $F(3, 156) = 1.44$ , *ns*,  $\eta_p^2 = .03$  or for target ethnicity,  $F(3, 154) = 0.78$ , *ns*,  $\eta_p^2 = .02$ , but we did find a significant interaction between rater ethnicity and target ethnicity,  $F(9, 374.9) = 2.08$ ,  $p < .05$ ,  $\eta_p^2 = .04$ .

To clarify the nature of the significant interaction we conducted four separate repeated measures ANOVA (one for each rater ethnic group) with physiological linkage as the dependent variable and target ethnicity as the repeated measure. Only the analysis for Chinese Americans yielded significant results,  $F(3, 40) = 2.76$ ,  $p = .05$ ,  $\eta_p^2 = .17$ . For Chinese Americans, the degree of physiological linkage during the rating task did depend on the ethnicity of the target being observed. Chinese American raters showed significantly greater physiological linkage when rating Chinese American targets ( $M = 33.1\%$ ,  $SD = 2.1$ ) than when rating European American targets ( $M = 25.6\%$ ,  $SD = 1.6$ ),  $t(42) = 2.94$ ,  $p < .01$ . A similar pattern was present for European American participants (greater linkage when rating European Americans), but this difference did not reach statistical significance,  $F(3, 35) = 1.57$ , *ns*,  $\eta_p^2 = .12$ .

#### Additional Analyses

To explore our finding that African American targets were rated most accurately, we conducted a post hoc analysis using the ratings provided by participants immediately after they rated each target. This took the form of repeated measure ANOVAs with target ethnicity serving as a repeated measure and ratings of how en-

<sup>3</sup> We thank a reviewer for suggesting these additional analyses regarding absolute versus cultural advantages and David Wagstaff for his statistical consultation in carrying them out.

<sup>4</sup> Degrees of freedom reflect adjustments for violations of underlying assumptions for a particular procedure.

Table 3  
Mean Accuracy Scores for the Entire Sample When Rating Each of Four Different Ethnic Targets

Accuracy scores	Target ethnicity			
	European American, <i>M (SE)</i>	Chinese American, <i>M (SE)</i>	African American, <i>M (SE)</i>	Mexican American, <i>M (SE)</i>
Lag zero				
Positive accuracy	1.89 (0.43) <sub>a</sub>	4.18 (0.71) <sub>b</sub>	6.00 (0.59) <sub>c</sub>	-0.25 (0.26) <sub>d</sub>
Negative accuracy	3.83 (0.61) <sub>a</sub>	5.08 (0.62) <sub>a</sub>	10.15 (0.88) <sub>b</sub>	1.38 (0.43) <sub>c</sub>
Lag one				
Positive accuracy	1.83 (0.44) <sub>a</sub>	4.04 (0.71) <sub>b</sub>	6.02 (0.61) <sub>c</sub>	-0.27 (0.23) <sub>d</sub>
Negative accuracy	3.77 (0.62) <sub>a</sub>	6.62 (0.75) <sub>b</sub>	9.59 (0.88) <sub>c</sub>	1.54 (0.42) <sub>d</sub>

Note.  $N = 161$ . Means in the same row that do not share the same subscript differ at  $p < .05$ , based on Fisher's least significant difference post hoc paired comparisons.

gaged they were by the conversation, difficulty of the rating task, perceived accuracy of ratings, and likability and attractiveness of targets as the dependent measure.<sup>5</sup>

The analyses revealed significant differences across the four target ethnicities in how engaged participants were with the conversation,  $F(3, 342) = 3.29, p < .05$ , how attractive they found the targets,  $F(3, 450) = 40.1, p < .001$ , and how likable they found the targets,  $F(2.54, 380.5) = 14.90, p < .001$ . Follow-up pairwise comparisons (Bonferroni adjusted) revealed that raters felt significantly less engaged by the African American target conversations relative to the Chinese American target conversations (see Table 5 for complete results). Analysis of the attractiveness and likability ratings for the four target ethnicities revealed that African American targets received significantly higher scores than European American and Mexican American targets, but did not differ from Chinese American targets. In sum, although participants were not more engaged in viewing the African American targets, they did report the African American targets to be more attractive and likable than their European American and Mexican American counterparts.

### Discussion

The goal of this study was to examine the effects of ethnic match on empathic accuracy and physiological linkage. Two models were considered, each of which clearly makes different predictions about the effects of ethnic similarity: (a) a cultural advantage model that predicts greater empathic accuracy and greater physiological linkage when individuals rate the emotions of members of their own ethnic/cultural group compared to a different ethnic/cultural group, and (b) a cultural equivalence model that predicts no cultural differences for empathic accuracy or physiological linkage regardless of whether raters rate the emotions of members of their own or a different ethnic/cultural group. Because previous studies have attempted to answer this question using static emotional stimuli (e.g., photographs) and one-time emotion judgments, the present study's use of dynamic, interpersonal emotional stimuli (conversations of couples) and continuous, real time emotion judgments offers an opportunity to revisit this question using stimuli and ratings that arguably more closely approximate the conditions under which these judgments are made in the real world.

### Empathic Accuracy

Our results for empathic accuracy revealed no support for the cultural advantage model. This was the case when we tested for an absolute cultural advantage as well as when we tested for the presence of a relative cultural advantage. These results are consistent with those of many previous studies in the emotion recognition literature that are more supportive of a cultural equivalence model of empathic accuracy (Boucher & Carlson, 1980; Gitter, Black, & Mostofsky, 1972a, 1972b; Gitter, Kozel, & Mostofsky, 1972; Kilbride & Yarczower, 1983; McAndrew, 1986; Sogon & Masutani, 1989). Our findings also are consistent with more recent studies that have attempted more stringent tests of the cultural advantage. For example, Beaupré and Hess (2005) failed to find a cultural advantage when a careful stimulus equivalence procedure was used to select target pictures. Similarly, Matsumoto, Ollide, and Willingham (2009) failed to find a cultural advantage when using spontaneous expressions of emotions as targets, arguably a more ecologically valid stimulus than posed expressions.

On the other hand, our findings are not consistent with work that demonstrates an in-group advantage in emotion recognition (e.g., Albas et al., 1976; Freeman, 1984; Weathers et al., 2002), nor with Elfenbein and Ambady's (2002b) meta-analytic review, which found support for an in-group advantage in the recognition of emotion across cultures. There are a number of possible explanations for why such an in-group advantage did not emerge in our study. As noted previously, we used a method for assessing empathic accuracy that attempted to maximize ecological validity. Most of the studies included in the Elfenbein and Ambady (2002b) review used a different kind of task in which participants observe a static emotional stimulus (usually a photo of a person portraying an emotional facial expression) and then apply the correct emotion term (a single emotion judgment). In our study, empathic accuracy was assessed in terms of ability to track the changing valence and intensity of a target's emotion during a 15-min conversation with a relationship partner. In this kind of real-life, dynamic, interpersonal context, the need to perceive another person (whether friend

<sup>5</sup> Rater ethnicity was not included as a between-subject variable because the overall MANOVAs for empathic accuracy revealed that there were no significant differences in accuracy across the four ethnic groups.

Table 4  
*Physiological Linkage Scores by Rater and Target Ethnicity*

Rater ethnicity	Target ethnicity (%)			
	European American, <i>M (SE)</i>	Chinese American, <i>M (SE)</i>	African American, <i>M (SE)</i>	Mexican American, <i>M (SE)</i>
European American <sup>a</sup>	32.0 (2.0) <sub>a</sub>	30.3 (1.9) <sub>a</sub>	28.9 (1.8) <sub>a</sub>	25.4 (2.3) <sub>a</sub>
Chinese American <sup>b</sup>	25.6 (1.6) <sub>a</sub>	33.1 (2.1) <sub>b</sub>	28.4 (2.0) <sub>a,b</sub>	28.9 (2.0) <sub>a,b</sub>
African American <sup>a</sup>	28.9 (2.1) <sub>a</sub>	24.8 (2.2) <sub>a</sub>	24.8 (2.1) <sub>a</sub>	27.4 (1.9) <sub>a</sub>
Mexican American <sup>c</sup>	24.6 (1.7) <sub>a</sub>	28.7 (2.0) <sub>a</sub>	29.1 (1.8) <sub>a</sub>	27.4 (1.6) <sub>a</sub>

*Note.* Means in the same row that do not share the same subscript differ at  $p < .05$ , based on Fisher's least significant difference post hoc paired comparisons.

<sup>a</sup>  $n = 38$ . <sup>b</sup>  $n = 43$ . <sup>c</sup>  $n = 41$ .

or foe) accurately may simply outweigh or overshadow any cultural advantage.

Another possibility is that we inadvertently stacked the deck against finding a cultural advantage in empathic accuracy by our choice of participants. We used selective eligibility criteria to ensure that selected participants had equal amounts of exposure to their cultures of origin and American mainstream culture. By employing these criteria we felt we could more confidently attribute any found differences to actual cultural differences. Nevertheless, all participants were recruited from the same geographic region. Research has demonstrated that part of the in-group advantage is diminished when in-group members have greater exposure to members of the out-group (Elfenbein & Ambady, 2003). Thus, a similar effect may have played out with the participants in our study whereby frequent contact with members of the other ethnic groups may have tempered a possible cultural advantage.

A third possibility is that our design was not powerful enough to detect an in-group advantage that did exist. Because the cultural equivalence model essentially proposes the null hypothesis of no difference, it is important to consider whether any lack of findings constitute the true absence of an effect or a problem with statistical power. In particular, we are concerned with the effect size and power estimates associated with the tests of the Rater Ethnicity (between-subject variable)  $\times$  Target Ethnicity (within-subject variable) interaction for empathic accuracy and physiological arousal. Based on estimates computed by our analysis software (SPSS, version 16), the effect size for the empathic accuracy interaction analysis was  $\eta_p^2 = .02$  (equivalent to  $r = .15$ ) with an

observed power of .83. This effect size falls in the lower range of the 95% confidence interval for the overall effect size of the in-group advantage (0.15, 0.34) reported by Elfenbein and Ambady (2002b) in their meta-analysis. Thus, according to these estimates it appears that we had sufficient power to detect even the smallest expected effect.

### *Physiological Linkage*

As noted earlier, we view physiological linkage as reflecting processes of emotional contagion/emotional empathy (i.e., feeling what another person is feeling). Prior research has shown that physiological linkage occurs both between (Levenson & Gottman, 1983) and within (Gottman & Levenson, 1985) spouses when they watch videotapes of their interactions and rate each other's emotions and that greater physiological linkage is associated with greater empathic accuracy when viewing and rating the negative emotions of strangers (Levenson & Ruef, 1992). Analysis of linkage data provided our only support for the cultural advantage model in that Chinese Americans showed greater physiological linkage when rating Chinese American targets. This finding is reminiscent of a prior finding that observing members of the in-group in an emotional situation produces similar emotions in the observer (Roberts & Levenson, 2006). However, we are not able to answer two important questions raised by these findings. First, we do not know why this finding was limited to Chinese American participants viewing Chinese American targets, although it is tempting to consider the well-documented collectivism in Chinese

Table 5  
*Mean Participant Ratings of How Engaged They Were by the Conversation Being Rated and How Attractive and Likeable They Found Each Target*

Ratings of targets	Target ethnicity			
	European American, <i>M (SE)</i>	Chinese American, <i>M (SE)</i>	African American, <i>M (SE)</i>	Mexican American, <i>M (SE)</i>
Engaged <sup>a</sup>	4.69 (0.19) <sub>a</sub>	5.15 (0.17) <sub>b</sub>	4.63 (0.19) <sub>a</sub>	4.47 (0.18) <sub>a</sub>
Attractive <sup>b</sup>	2.89 (0.15) <sub>a</sub>	4.13 (0.15) <sub>b,c</sub>	4.36 (0.16) <sub>b</sub>	3.98 (0.14) <sub>c</sub>
Likeable <sup>b</sup>	4.44 (0.16) <sub>a</sub>	5.28 (0.12) <sub>b</sub>	5.48 (0.12) <sub>b</sub>	4.80 (0.13) <sub>a</sub>

*Note.* Means in the same row that do not share the same subscript differ at  $p < .05$ , based on Fisher's least significant difference post hoc paired comparisons.

<sup>a</sup>  $n = 115$ . <sup>b</sup>  $n = 151$ .

American culture (Hofstede, 2001) as a possible explanation. Second, we did not find greater physiological linkage in Chinese Americans to be associated with an in-group advantage in empathic accuracy as we would have predicted. Thus, replication of this finding using experimental designs that would enable us to address these questions is clearly called for.

### *Empathic Accuracy and Emotional Expression*

Although ethnic match did not emerge as an important determinant of how well we read emotions from others, we did find some evidence of cultural variation in empathic accuracy. Participants were consistently most accurate when rating the emotions of African Americans, followed by Chinese Americans, European Americans, and Mexican Americans. This finding is noteworthy because the conversations were selected in a manner that maximized their equivalence in terms of topic, amount of negative and positive affect, and ratio of negative affect to positive affect. However, differences in emotional expressiveness may still have affected the effectiveness with which emotion was conveyed.

Evidence from Johnson (1989) might begin to explain these findings. Johnson found that African Americans reported experiencing anxiety more frequently and anger more intensely than European Americans. Assuming that experiencing an emotion more frequently and more intensely is ultimately tied to more visible displays of the emotion, it is reasonable to expect that African Americans may be more demonstrative with at least some negative emotions. Given the explicitly negative focus of at least half of the stimulus materials, African American targets may have manifested greater and more detectable negative emotion thereby giving participants an edge in rating their emotions overall.

Also striking among our results is the level of empathic accuracy among participants when rating Chinese American and Mexican American targets. Previous research and ethnographic accounts suggest that Chinese culture embraces emotional moderation and that Mexican culture embraces emotional expression (Murillo, 1976; Potter, 1988; Soto et al., 2005). Based on this we might have expected Mexican American targets to be rated most accurately and Chinese American targets to be rated least accurately. However, our findings were almost exactly the opposite, with Mexican American targets being rated least accurately and Chinese American targets being rated second-most accurately. These results were surprising, but may be understood in light of results from our previous work examining patterns of emotional expression and experience among Chinese American and Mexican American college students exposed to an acoustic startle (Soto et al., 2005). We found that although Chinese Americans consistently reported feeling less emotion than their Mexican American counterparts, there were no differences in how much emotional behavior the two groups displayed. Thus, the notion that Chinese Americans are not very emotionally expressive and that Mexican Americans are, may only hold true when considering self-reports of emotion and not other channels of emotion such as observable behavior or physiology.

### *Strengths and Limitations*

As noted earlier, this study was designed to maximize ecological validity of the empathic accuracy assessment by using a task that

more closely resembles the way emotions are communicated in real life (real-time judgments from naturalistic behavior). We also aimed to use a participant recruitment strategy that ensured that participants had adequate exposure to the cultural traditions and other members of the in-group. These strengths notwithstanding, there were also several limitations of the study. These include: (a) the use of college students who may have extensive exposure and acculturation to the out-group cultures, thus lessening a possible in-group advantage; (b) using an empathic accuracy assessment based on assessment of emotional valence and intensity and not discrete emotions (as in previous photo identification studies); (c) not assessing similarity between raters and targets in expressive behavioral responses (only physiological linkage was assessed); and (d) variability related to ethnic groups in targets and conversations (the downside to using naturalistic stimuli).

The last of these points is arguably the most significant limitation of the present study: the potential lack of equivalency among the stimulus conversations. Ideally, the stimulus conversations used would have differed only in the ethnicity of the targets (see Beaupre & Hess, 2005; Matsumoto, 2002). Despite the fact that each conversation selected for the study met exacting selection criteria, there was no way to ensure that the conversations representing each ethnic group were equivalent in all important ways. Given the focus on culture in the present study, two representative stimuli from each cultural group were used. However, due to time constraints and considerations of participant fatigue, each participant only rated one conversation from each ethnic group. A more optimal design would have been to have each participant rate more than one representative from each cultural group, perhaps by using fewer ethnic groups and/or shorter conversations. Devising larger sets of well-matched stimulus conversations will be important for conducting future studies of culture and empathic accuracy.

### Conclusions

The question of how culture influences the communication of emotions has a long and rich history in the empirical literature. Evidence has been found in support of both the culture equivalence model of emotion recognition (Boucher & Carlson, 1980; Gitter, Kozel, & Mostofsky, 1972; Kilbride & Yarczower, 1983) and the cultural advantage model (Elfenbein & Ambady, 2002a, 2002b). The present study revisited this issue examining empathic accuracy and physiological linkage using a methodology that more closely approximates the ways that emotions are typically detected in everyday life. Our findings lent no support to the cultural advantage model in empathic accuracy, but did find some support for this model in physiological linkage for Chinese American participants and targets. Extending this research to other samples and other materials would help map out the contexts in which cultural equivalence and cultural advantage are found. The present results suggest that when making dynamic, real-time, judgments of others' emotions in the kinds of interpersonal contexts in which most such judgments occur, we are equally adept at recognizing the emotions of in-group and out-group members. This finding can be construed as hopeful news for increasing cooperation and communication in our increasingly multiethnic environments.

## References

- 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.
- Anderson, C., & Keltner, D. (2002). Open peer commentary: The role of empathy in the formation and maintenance of social bonds. *Behavioral and Brain Sciences, 25*(1), 21–22.
- Beaupré, M. G., & Hess, U. (2006). An ingroup advantage for confidence in emotion recognition judgments: The moderating effect of familiarity with the expressions of outgroup members. *Personality and Social Psychology Bulletin, 32*, 16–26.
- Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology, 11*, 263–280.
- Chevalier-Skolnikoff, S. (1974). *The ontogeny of communication in the stump-tail macaque* (Macaca arctoides). Oxford, UK: Kragar.
- Darwin, C. (1872/1965). *The expression of the emotions in man and animals*. Chicago: The University of Chicago Press. (Originally published, 1872).
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Boulder, CO: Westview Press.
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*, 406–412.
- Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *The Scientific World Journal, 6*, 1146–1163.
- Decety, J., & Meyer, M. (2008). From emotion resonance to empathic understanding: A social developmental neuroscience account. *Development and Psychopathology, 20*, 1053–1080.
- Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*, 643–647.
- Eisenberg, N., Fabes, R. A., Miller, P. A., Fultz, J., Shell, R., Mathy, R. M., et al. (1989). Relation of sympathy and personal distress to prosocial behavior: A multimethod study. *Journal of Personality and Social Psychology, 57*, 55–66.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation*, 1971 (pp. 207–283). 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., 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., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*, 1208–1210.
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*, 86–88.
- 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*, 203–235.
- Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: Cultural exposure and facial emotion recognition. *Journal of Personality and Social Psychology, 85*, 276–290.
- Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7*, 131–146.
- Freeman, E. B. (1984). The development of empathy in young children: In search of a definition. *Child Study Journal, 13*, 235–245.
- Gitter, A. G., Black, H., & Mostofsky, D. (1972a). Race and sex in the communication of emotion. *Journal of Social Psychology, 88*, 273–276.
- Gitter, A. G., Black, H., & Mostofsky, D. (1972b). Race and sex in the perception of emotion. *Journal of Social Issues, 28*(4), 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.
- Gottman, J. M. (1981). *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge, UK: Cambridge University Press.
- Gottman, J. M., & Levenson, R. W. (1985). A valid procedure for obtaining self-report of affect in marital interaction. *Journal of Consulting and Clinical Psychology, 53*, 151–160.
- Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 97–117). New York: Cambridge University Press.
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. New York: Cambridge University Press.
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.
- Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*, 587–610.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*, 288–299.
- Johnson, E. H. (1989). Cardiovascular reactivity, emotional factors, and home blood pressures in Black males with and without a parental history of hypertension. *Psychosomatic Medicine, 51*, 390–403.
- Kilbride, J. E., & Yarczower, M. (1983). Ethnic bias in the recognition of facial expressions. *Journal of Nonverbal Behavior, 8*(1), 27–41.
- Klein, K. J. K., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*, 720–730.
- LaFrance, M., & Banaji, M. (1992). Toward a reconsideration of the gender-emotion relationship. In M. S. Clark (Ed.), *Emotion and social behavior* (pp. 178–201). Thousand Oaks, CA: Sage.
- Levenson, R. W. (2003). Autonomic specificity in emotion. In R. Davidson, K. Scherer, & H. Goldsmith (Eds.), *Handbook of affective science* (pp. 212–224). New York: Oxford University Press.
- Levenson, R. W., & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology, 45*, 587–597.
- Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234–246.
- Levenson, R. W., & Ruef, A. M. (1997). Physiological aspects of emotional knowledge and rapport. In W. J. Ickes (Ed.), *Empathic accuracy* (pp. 44–72). New York: Guilford Press.
- Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences, USA, 106*, 9925–9930.
- Linnankoski, I., Laakso, M.-L., Aulanko, R., & Leinonen, L. (1994). Recognition of emotions in macaque vocalizations by children and adults. *Language and Communication, 14*, 183–192.
- Markham, R., & Wang, L. (1996). Recognition of emotion by Chinese and Australian children. *Journal of Cross-Cultural Psychology, 27*, 616–643.
- Marsh, A. A., Elfenbein, H. A., & Ambady, N. (2003). Nonverbal "accents": Cultural differences in facial expressions of emotion. *Psychological Science, 14*, 373–376.
- 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. (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.
- Matsumoto, D. (2007). Apples and oranges: Methodological requirements for testing a possible ingroup advantage in emotion judgments from facial expressions. In U. Hess (Ed.), *Group dynamics and emotional expression* (pp. 140–181). New York: Cambridge University Press.
- Matsumoto, D., Olide, A., & Willingham, B. (2009). Is there an ingroup advantage in recognizing spontaneously expressed emotions? *Journal of Nonverbal Behavior*, *33*, 149–211.
- McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *17*, 211–224.
- Murillo, N. (1976). The Mexican American family. In C. A. Hernandez, M. J. Haug, & N. N. Wagner (Eds.), *Chicanos: Social and psychological perspectives* (pp. 97–108). St Louis, MO: Mosby.
- O'Toole, A. J., Peterson, J., & Deffenbacher, K. A. (1996). An "other-race effect" for categorizing faces by sex. *Perception*, *25*, 669–676.
- Potter, S. H. (1988). The cultural construction of emotion in rural Chinese social life. *Ethos*, *16*, 181–208.
- Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*, 1–72.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
- Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1999). Resonance behaviors and mirror neurons. *Archives Italiennes De Biologie*, *137*, 85–100.
- Roberts, N. A., & Levenson, R. W. (2006). Subjective, behavioral, and physiological reactivity to ethnically matched and ethnically mismatched film clips. *Emotion*, *6*, 635–646.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, *115*, 102–141.
- Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin*, *118*, 379–391.
- Sogon, S., & Masutani, M. (1989). Identification of emotion from body movements: A cross-cultural study of Americans and Japanese. *Psychological Reports*, *65*, 35–46.
- Soto, J. A., Levenson, R. W., & Ebling, R. (2005). Cultures of moderation and expression: Emotional experience, behavior, and physiology in Chinese Americans and Mexican Americans. *Emotion*, *5*, 154–165.
- Soto, J. A., Pole, N., McCarter, L. M., & Levenson, R. W. (1998, September). *Knowing feelings and feeling feelings: Are they connected?* Paper presented at the 38th annual meeting of the Society for Psychophysiological Research, Denver, CO.
- Thibault, P., Bourgeois, P., & Hess, U. (2006). The effect of group-identification on emotion recognition: The case of cats and basketball players. *Journal of Experimental Social Psychology*, *42*, 676–683.
- Vinacke, W. E., & Fong, R. W. (1955). The judgment of facial expressions by three national-racial groups in Hawaii: II. Oriental faces. *Journal of Social Psychology*, *41*, 185–195.
- Weathers, M. D., Frank, E. M., & Spell, L. A. (2002). Differences in the communication of affect: Members of the same race versus members of a different race. *Journal of Black Psychology*, *28*(1), 66–77.
- 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.
- Zaki, J., Weber, J., Bolger, N., & Ochsner, K. N. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences, USA*, *106*, 11382–11387.

Received September 29, 2008

Revision received July 15, 2009

Accepted August 3, 2009 ■