

American-Japanese cultural differences in judgements of emotional expressions of different intensities

David Matsumoto and Theodora Consolacion

San Francisco State University, USA

Hiroshi Yamada and Ryuta Suzuki

Nihon University, Japan

Brenda Franklin, Sunita Paul, Rebecca Ray, and Hideko Uchida

San Francisco State University, USA

Although research has generated a wealth of information on cultural influences on emotion judgements, the information we have to date is limited in several ways. This study extends this literature in two ways, first by obtaining judgements from people in two cultures of expressions portrayed at different intensity levels, and second by incorporating individual level measures of culture to examine their contribution to observed differences. When judging emotion categories in low intensity expressions, American and Japanese judges see the emotion intended at above-chance rates, albeit lower than when judging high intensity faces. Also, American and Japanese intensity ratings of external displays and internal experiences differ dramatically for low intensity expressions compared to high intensity faces. Finally, the two cultural dimensions measured in this study—individualism versus collectivism (IC) and status differentiation (SD)—accounted for almost all of the variance in the observed differences. These findings are discussed in terms of their underlying possible mechanisms, and future research possibilities.

Research over the past three decades has shown ample evidence of pancultural agreement in judgements of emotion categories from facial expressions in literate (Ekman, Sorenson, & Friesen, 1969; Izard, 1971) and preliterate cultures (Ekman & Friesen, 1971; Ekman et al., 1969). These findings replicate when observers judge spontaneous expressions (Ekman, 1972), and the presence,

Correspondence should be addressed to David Matsumoto, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, USA; e-mail: dm@sfsu.edu

We thank Haruyo Tatani for her assistance in our general research program.

absence, and intensity of multiple emotions (Ekman et al., 1987). Cultures also agree on their second modal judgements, and relative intensity differences within emotions (Ekman et al., 1987; Matsumoto & Ekman, 1989).

Cultures differ, however, in the exact level of agreement in categorical judgements (Matsumoto, 1992; Russell, 1994; see rebuttals to Russell, 1994 by Ekman, 1994, and Izard, 1994), and those differences are related to stable and meaningful dimensions of cultural variability (Matsumoto, 1989). Cultures also differ in judgements of intensity (Biehl et al., 1997; Ekman et al., 1987; Matsumoto, 1990), regardless of the poser's race or sex (Matsumoto, 1990; Matsumoto & Ekman, 1989) and across ethnicities (Matsumoto, 1993). In these studies, Americans have rated expressions more intensely than Asians, particularly Japanese. But this effect is limited to ratings of external display; when rating presumed internal experience, Japanese give higher ratings than Americans (Matsumoto, Kasri, & Kookan, 1999). These effects occurred because Americans rated external displays higher than internal experience, while there was no difference between the two ratings for Japanese.

These studies raise important theoretical and methodological questions that need to be addressed in order for contemporary cross-cultural research on emotion judgements to advance. This study does so in two arenas, one concerning the nature of the expressions used as stimuli, the second concerning the definition and measurement of culture.

Expressions of varying intensities

One common feature in studies in this area is the use of full-face, high intensity expressions as stimuli. To be sure, these stimuli are useful when testing for universality. Their exclusive use is limiting, however, especially because data indicating the incidence of some full-face, prototypic expressions (e.g., anger, fear, sadness) in real life are sparse, and because many naturally occurring expressions are subtle, often laden with cultural meaning and influenced by cultural display rules.

One way to address this limitation is to include different types of faces in research. For example, expressions can be shown at different intensity levels, as blends of two or more emotions, as partial emotions (i.e., portrayed in one area of the face but not another), or as profiles. Ultimately, spontaneous expressions systematically varied in terms of their muscle dynamics—onset, apex, offset, and laterality—portrayed in context could also be used.

Hess, Blairy, and Kleck (1997) recently created expressions differing in intensity levels using the JACFEE expressions used in this study (see Methods). Twenty-four observers from an American university judged expressions at 0%,

20%, 40%, 60%, 80%, and 100% increments, by designating full-face JACFEE expressions as 100% and their corresponding neutral expressions as 0%, and morphing them to produce expressions at the remaining intensity intervals. They found that both intensity ratings and accuracy judgements of the expressions increased in a linear fashion as expression intensity increased.

In this study, we also created stimuli portraying emotion at different intensity levels. The availability of neutral and full-faced expressions of the same poser, and of computer morphing technology, allowed us to create “Low” and “Very High” intensity expressions of full-face (see Methods for details), designating the neutral and original expressions as “Neutral” and “High” intensities, respectively (see Figure 1 for examples). These stimuli allowed us to address important questions about the cross-cultural nature of emotion judgments that have been unaddressed previously. While agreement among observers increased as expression intensity increased in Hess et al.’s (1997) study, we do not know whether this is true across cultures or on a categorical judgement task.¹ This study will be the first to address this issue.

Differences in expression intensity (and thus, signal clarity) may also have consequences for cultural differences. For example, do the differences reported in the literature (e.g., Matsumoto, 1989, 1992) in categorical judgements exist for subtler, or more intense, expressions as well? On one hand, cultural differences may be eliminated because of the increased ambiguity of low intensity stimuli. On the other hand, this increased ambiguity may exacerbate cultural differences, leading to more pronounced differences than with full-face expressions, because judges will need to “read into” the faces more, and in doing so, utilize cultural biases.

The inclusion of very high intensity expressions extends previous research and knowledge. Although the full-face expressions used in most previous cross-cultural studies do indeed present expressions at relatively high intensity levels,

¹In Hess et al.’s (1997) study, observers rated the intensity of each expression using multiple emotion scales. Accuracy was defined as whether or not an observer gave the intended emotion scale the highest intensity rating, and coding those expressions as “1”; all other expressions were then coded “0”. The linear trends reported by Hess et al. (1997) with regard to their accuracy data may have been influenced by the scalar nature of the data obtained. We (Matsumoto et al., 1999) have argued elsewhere that single emotion intensity ratings are ambiguous as to exactly what is being rated—the intensity of the expression displayed, or the intensity of the presumed subjective experience. Previous cross-cultural research has indeed produced cultural differences in opposite directions, depending on which intensity rating is considered (Matsumoto et al., 1999). In this study, we address this ambiguity by obtaining a categorical judgement of the expression, an intensity rating of the expression, and an intensity rating of the presumed subjective experience (see Methods for details).

Neutral



Low



High



Very High

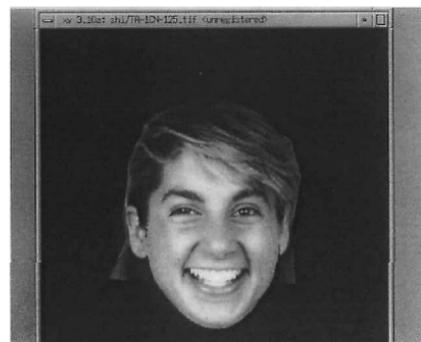


Figure 1. An example of the stimuli used in this study.

they generally do not represent expressions at their greatest intensity.² Five-point (A–E) intensity ratings using Ekman and Friesen’s Facial Action Coding System (FACS) of the JACFEE expressions used in this study (as well as Hess et al., 1997; see Methods), for example, typically involve intensity ratings at the 3rd and 4th levels (FACS levels C and D). More intense expressions do occur in real life (FACS level E intensity), and yet have never been studied in judgement research in the past. Thus, many of the questions raised about low and high intensity expressions, including levels of agreement, incremental increases in agreement, and judgements of intensity, can be raised about very high intensity expressions as well. Their inclusion in this study allows us to examine whether or not judgements of these expressions differ in a linear fashion across expression intensities.

The stimuli also allowed us to explore further the nature of previously reported American-Japanese cultural differences in intensity ratings. Does this effect replicate with less intense and very intense, expressions? Americans may exaggerate their ratings of external display relative to internal experience for all expressions, regardless of intensity. Or, they may do so only with high intensity expressions. Japanese may continue to not differentiate the intensity between external display and internal experience, regardless of expression intensity. Or, low intensity expressions may be interpreted as suppressed and controlled, in congruence with assumed Japanese cultural display rules, resulting in higher ratings of presumed internal experience relative to external display.

Unpackaging the effects of culture

The second arena in which this study makes a unique contribution is its exploration of the specific nature of cultural influences on judgments. Most cross-cultural work in this area until now has really been cross-national, with culture operationalised by country. Although this approach has been standard practice, it limits our ability to interpret differences when observed. That is, when culture is operationalised by country and differences are found, they can only be interpreted by making assumptions about the nature of cultural differences underlying the countries, because culture was never actually measured.

Recently, several writers have called for a move away from such practices, and towards research that “unpackages” the culture effects on psychological variables (e.g., Bond & Tedeschi, 2001; Poortinga, Van de Vijver, Joe, & Van de Koppel, 1987; Van de Vijver & Leung, 1997). Unpackaging refers to the

² And, these same full-face expressions used in previous studies were the expressions displayed at greatest intensity (100%) in Hess et al.’s (1997) study.

identification of specific, psychological dimensions related to culture³ that may account for between-country differences in the variable of interest, their inclusion and measurement, and the statistical estimation of the degree to which they actually account for between-country differences. Thus, specific, measurable dimensions related to culture on the psychological level replace the global, nonspecific construct we know of as “culture”.

Individualism versus collectivism. One of the most important and widely used cultural dimensions is known as individualism versus collectivism (IC). Individualistic cultures tend to emphasize the individual, fostering uniqueness, separateness, autonomy, and individuality. Collectivistic cultures value ingroups over individuality, and foster harmony, cohesion, and cooperation. IC has been used to explain many cross-national and cross-cultural differences in behaviour, and is arguably the most well-known, well-studied, and important dimension of culture that exists today (Triandis, 1994, 1995).

The field has made considerable inroads into the creation of valid and reliable methods of measuring IC-related attitudes, values, and beliefs on the individual level, most of it led by Triandis and his colleagues. On the individual level, IC tendencies are referred to as idiocentrism and allocentrism, respectively (Triandis, Leung, Villareal, & Clack, 1985). Triandis (1995) views IC as a syndrome, cutting across a variety of psychological domains; consequently, his multimethod assessment technique measures IC tendencies across these domains. Other techniques also exist, such as Hui's (1984, 1988) context-specific measure of collectivism, Yamaguchi's (1994) multidomain measurement of collectivism, and Singelis' (Singelis, Triandis, Bhawuk, & Gelfand, 1995) measure of horizontal and vertical dimensions of IC.

³ It is important to note that there is a difference between individual-level measures of psychological dimensions related to culture, and culture itself. The former refers to measurable psychological domains such as attitudes, values, beliefs, and opinions that are either theoretically or empirically related to sociocultural dimensions that differentiate among cultures, such as Hofstede's (1980, 1983) dimensions of Individualism, Power Distance, Uncertainty Avoidance, and Masculinity. Indeed, there is a controversy in the field concerning the ability of such individual-based measures to capture cultural differences between groups. On one hand, Triandis and his colleagues (summarised in 1994, 1995) would suggest that their individual-based measures of I-C do capture group-level cultural phenomena. On the other hand, Peng, Nisbett, and Wong (1997) would argue that the implicit social comparisons respondents make when completing individual-level value surveys may invalidate between group differences in those same surveys. The current study is not designed to address this important issue in the field; instead, we aim to examine the degree to which individual differences on culture-related psychological dimensions can account for between-country differences in emotion judgements, without making the assumption that those individual differences measure the larger, more global construct of culture itself. Consequently, we are careful in the remainder of the article in our language concerning the variables used to unpackage culture, referring to them as individual-level measures of constructs *related* to culture.

Matsumoto has also developed a measure of IC specific to social interaction (Matsumoto, 1997), which was used in the present study. Termed the Individualism and Collectivism Interpersonal Assessment Inventory (ICIAI), participants rate 19 items reflecting the degree of importance for themselves of IC-related behavioural, attitudinal, and value tendencies in four social relationships: family members, close friends, work colleagues, and strangers in two ratings domains (values and behaviours) using 7-point scales. (The scales are anchored 0, Not at All Important, to 6, Very Important for the values ratings, and 0, Never, to 6, All the Time, for behaviours.) Four scales can be scored:⁴ Social Harmony (SH), Social Identification (SI), Self-Control (SC), and Social Sharing of Recognition (SSR) across relationships, and for each relationship across scales. Differences among the scales reflect different aspects of individual-level IC, while differences among the relationships reflect differential IC-mediated self-ingroup and outgroup tendencies (see Triandis, 1988). A total score averaging across all scales and relationships is also computed.

As a global, cultural construct, IC is important to consider in explaining cross-national differences in emotion judgements for several reasons. It has been linked with cultural differences in a wide variety of emotion-related phenomena, including judgements, experience, and display rules (e.g., see Matsumoto, 1989; Matsumoto, Takeuchi, Andayani, Kouznetsova, & Krupp, 1998; Wallbott & Scherer, 1986). Some research (e.g., Matsumoto, 1990, 1991) has suggested that people of individualistic cultures express feelings more openly than people of collectivistic cultures. This suggests that people of collectivistic cultures may ‘read into’ others’ expressions more, assuming that people are actually feeling more than they show, because they recognise the display rule to deamplify expressions relative to experience. This also suggests that people of individualistic cultures may presume that others are not feeling their emotions as much as they show it, compensating for their display rules. This would explain Matsumoto et al.’s (1999) report of American-Japanese cultural differences in intensity ratings, where Americans gave higher ratings to external displays than to attributed internal experience, whereas there was no difference between the two ratings for the Japanese.⁵

This reasoning also leads to another hypothesis concerning the accuracy of emotion judgements. If Americans and Japanese take supposed display rules into

⁴ The scales and their computational methods are based on pancultural factor analyses conducted on normative data collected in countries, including the United States and Japan.

⁵ That Americans give higher ratings to external display than to the poser’s presumed internal experience suggests that they believe posers are not feeling emotions to the same degree as they are expressing them. This, in turn, implies the existence of a display rule to express more than is felt. In that study (Matsumoto et al., 1999), however, display rules were not measured; thus, the portion of this interpretation involving the supposed working of display rules remains speculative, and is discussed further in the discussion.

account when making judgements of others, it would not be unreasonable to suspect that Japanese would be better able to identify emotions when the signal to noise ratio is low.⁶ That is, given greater ambiguity in a stimulus, the Japanese may be more likely to attribute the intended emotion with less cues, as they may be compensating for the lack of cues in their judgements. That previous research (Matsumoto, 1989, 1992) testing American-Japanese agreement levels in judgements of emotion categories has typically shown that Americans have higher levels of agreement may not be a factor in this rationale, as those studies have used full-face, high intensity stimuli that are quite nonambiguous. When more ambiguous, low intensity expressions are used, such as in this study, the Japanese may be more accurate in their judgements of emotion than Americans.

Status differentiation. In addition to IC, other important psychological dimensions of culture exist, such as power distance, status differentiation, uncertainty avoidance, masculinity, tightness, and contextualisation (Hall, 1996; Hofstede, 1980, 1983; Matsumoto, 1991; Pelto, 1968). One issue that has hampered research has been the unavailability of valid and reliable measures of them. Recently, Matsumoto and his colleagues developed an individual difference measure of status differentiation (Kooken, Yrizarry, Consolacion, Vogt, & Matsumoto, 1997; Kooken, Yrizarry, & Matsumoto, 1996; Seki, Yrizarry, Kooken, & Matsumoto, 1995), which refers to the degree to which cultures differentiate their behaviours toward others on the basis of the status differences that exist between them and their interactants. Some cultures make large differentiations, affording people of higher status more power (i.e., the ability to influence others' behaviours); others make smaller differentiations, treating people more or less the same regardless of status differences.

In Matsumoto's Status Differentiation Scale (SDS), which was used in this study, respondents rate 20 items reflecting attitudinal and behavioral tendencies on their appropriateness in relation to interactions with people at three status levels (same, higher, lower) in two domains—work and school. Ratings are made on 7-point scales, anchored 0, Not Appropriate at All, to 6, Very Appropriate. Scores are computed by summing the absolute difference between same versus higher and same versus lower status across all items, separately for each domain. Scores are also computed for each status difference across domain, and for each domain across status difference. A total score is also computed.

SD is also important to consider in cross-cultural work on emotion because, among other reasons, IC never accounts for the entire effect size associated with cross-national differences. In considering other dimensions that may account for such differences, it is reasonable that the search include differences in the use of status and power, as many writers have alluded to the importance of these

⁶This hypothesis was suggested by one of the reviewers.

concepts (e.g., Hofstede, 1980, 1983; Mulder, Veen, Hijzen, & Jansen, 1973; Mulder, Veen, Rodenburg, Frenken, & Tielens, 1973). Also, some research has suggested that people differentiate their emotional displays according to the status differential of others with whom they interact (Matsumoto, 1990, 1991). People of high status differentiating cultures tend to smile more and show negative emotions less toward higher status others than people of low status differentiating cultures. They also tend to smile less and show more negative emotions to lower status others. If true, people of higher status differentiating cultures may read more into others' expressions, recognising the possible influence of SD-influenced cultural display rules on expression.

Individual-level measurement techniques of dimensions related to constructs such as IC or SD are major pluses for cross-cultural research because they allow researchers to examine empirically the degree to which individual-level assessment of culture-related attitudes, values, behaviours, and opinions account for group-level differences in psychological phenomena. This examination can be done without making any a priori assumptions about whether or not such measures actually capture the larger, more global constructs of culture that purportedly underlie between country differences. How much between-country differences can be accounted for by individual-level measures of culture-related psychological variables is an empirical question that can and should be addressed irrespective of whether such variables *are* culture. Although IC, and to a lesser extent SD, has been used to explain differences in a variety of phenomena, including emotion and judgements, no study has examined exactly how much individual-level attitudes, values, and beliefs concerning IC and SD contribute to these variables. This is addressed in this study.

An example of the quantification of the contribution of a specific individual-level cultural variable's to cross-national differences. Although several alternatives exist, the procedures adopted in this study mirror that used in Matsumoto et al. (1998), in which emotion display rules were assessed in respondents from four countries, who also completed Matsumoto et al.'s (1997) ICIAI. Cross-national differences in the display rules were assessed through the usual ANOVA procedures, and their effect sizes were computed. Then, group differences were re-tested using ICIAI scores as covariates, and new country effect sizes were computed. Matsumoto et al. (1998) interpreted the difference between the original country effect size and the new effect size to reflect the contribution of the covariate to the cross-national difference. A ratio of this difference to the original effect size thus reflects the proportion of the original effect size accounted for by the covariate. In their study, individual-level IC-related attitudes, values, and beliefs generally accounted for approximately 30% of the observed cross-national differences in display rules.

Overview of this study and hypotheses

Using expressions portraying emotion at multiple intensity levels, we asked observers in two countries—US and Japan—to make three judgements of each: a categorical emotion judgement, and intensity ratings of the external display of the expression and the presumed internal experience of the poser. Prior to the judgement tasks, all respondents completed measures of IC and SD. On the basis of previous research, and the rationales provided above, we tested the following hypotheses with regard to the categorical judgements:

Hypothesis 1: That observers will judge low intensity expressions as portraying the emotions intended at levels greater than chance, and greater than that for neutral expressions; but

Hypothesis 2: That the agreement levels for low intensity expressions will be less than for high intensity expressions; and

Hypothesis 3: That Japanese will have higher agreement levels than Americans for low intensity expressions.

On the basis of previous research, we tested the following hypothesis regarding the two intensity ratings:

Hypothesis 4: That Americans will give significantly higher ratings to external display relative to internal experience for low intensity expressions, while there would be no difference between the two ratings for the Japanese.

Finally, the inclusion of the two individual-level cultural variables allowed us to “unpackage” culture and its effects on the observed judgements and to test the following hypothesis:

Hypothesis 5: That individual-level IC and SD account for a substantial proportion of the between-country variance in judgements of external display and internal experience.

METHODS

Participants

A total of 275 undergraduates were recruited from psychology courses and participated as noncompensated volunteers. Of these, 138 were Americans (94 females, 44 males, mean age = 24.93, SD = 5.31) while 137 were Japanese (74 females, 63 males, mean age = 20.20, SD = 1.29). (The higher age for the Americans reflected the fact that the university from which they were recruited included a large proportion of commuter and returning students. Age was correlated with the emotion judgement variables reported below, but no effects were found.) All participants and their parents were born and raised in their

respective country; their native languages were English and Japanese, respectively. The majority of both samples reported their socioeconomic status as middle income, and the bulk was juniors and seniors.

Facial stimuli

The stimuli included 64 expressions adapted from Matsumoto and Ekman's (1988) Japanese and Caucasian Facial Expressions of Emotion (JACFEE). This widely used stimulus set reliably portrays the universal emotions. Four posers were randomly selected from the angry, happy, sad, and surprised expressions, and matched with their neutral expression from Matsumoto and Ekman's (1988) Japanese and Caucasian Neutral Faces (JACNeuF). A total of 739 reference points were placed onto a digitised version of each JACFEE high intensity expression and its corresponding neutral. Of these, 28 were manually chosen on the outline of the face, 8 on the eyes, 8 on the brows, 4 on the nose, 6 on the mouth, and 15 on the hair; the remainder was chosen randomly. With the JACFEE expression set at 100% and the neutral set at 0%, the low and very high intensity expressions were generated by producing expressions with their reference points at 50% and 125%, respectively, of the distance between the JACFEE neutral and high intensity expression. The final stimulus set included 64 expressions – 4 emotions \times 4 posers \times 4 intensity levels). In addition, three JACFEE posers that were not utilised for morphing were used for practice trials. All stimuli were inspected by coders trained in Ekman and Friesen's (1978) Facial Action Coding System to ensure that they portrayed expressions that were anatomically possible, involving the same action units as the original expressions. No discrepancies were found.

Judgment task

Participants made three ratings for each expression. The first was a categorical judgement of the emotion they thought best described the expression, selecting a single choice from nine alternatives: anger, contempt, disgust, fear, happiness, sadness, surprise, no emotion, and other. For "other" responses, a blank line was provided to allow open-ended responses. Participants also rated the intensity of the external display of the expression, and the subjective experience of the poser, using two 9-point scales labelled None, 0, to A Lot, 8. The instructions for these ratings were exactly the same as those used in Matsumoto et al. (1999), and were as follows:

Your task is to: (1) determine what emotion if any is being displayed and to check the corresponding box using the scale provided. Please check only one box. If the emotion is not listed or if you believe multiple emotions are present, please write

what you believe is being displayed on the blank line marked "Other". (2) Indicate the intensity level of that facial expression and (3) indicate the intensity level of what you think the poser is actually feeling using this 0 to 8 scale. Please choose only the numbers used on this scale.

Individual-level measures of culture-related dimensions

Participants completed Matsumoto et al.'s (1997) ICIAI and SDS (described earlier). Cronbach's alpha for both scales for both countries indicated a high degree of internal reliability: .92 and .93 for the ICIAI, and .91 and .88 for the SDS for the United States and Japan, respectively. Correlational analyses on both the ICIAI and SDS indicated substantial overlap between the total scores and all other derived scores; thus, only the total scores for both measures were used in the main analyses reported below.

Procedures

All participants were tested in small groups. On arrival at the laboratory, participants completed a comprehensive demographics sheet providing personal information, and then the ICIAI and SDS, which took approximately 20 minutes. The experimenter then explained the judgement task and answered all questions. The three practice expressions were shown to the participants, who made their three judgements on each. Any further questions were answered. There were no confusions about the meaning or intent of the procedures. The 64 pictures were divided into four groups of 16 so that no poser appeared more than once in a group, and each group contained an equal number of emotions and expression intensities. Order within each group was determined randomly, and participants were randomly assigned to view one of the four groups of stimuli. The experimenter then showed the expressions, one at a time, to the group, for approximately 10 seconds each, via a large screen computer monitor in ample view of all participants. Completion of the ratings of all 16 expressions marked the end of the experiment.

RESULTS

Categorical judgements: Hypotheses 1–3

Main analyses. The percentage of observers in both countries selecting the intended emotion category for each expression is given in Table 1. We tested whether these percentages were significantly greater than chance (1/9) using chi-

square, separately for the two countries.⁷ For low intensity expressions, 30 out of 32 comparisons (16 expressions \times 2 countries) were statistically significant, all in the direction predicted. With the exception of happiness, the percentage of observers choosing the intended emotion category for neutral expressions was negligible, as would be expected, and 30 of the 32 comparisons of neutral versus low expressions indicated that a greater percentage of observers chose the intended emotion category for low intensity expressions. Collectively, these findings strongly supported Hypothesis 1, which stated that observers will judge low intensity expressions as portraying the emotions intended at levels greater than chance, and greater than that for neutral expressions.

To test Hypothesis 2—that the agreement levels for low intensity expressions will be less than for high intensity expressions—we computed chi-squares comparing the percentage of respondents selecting the intended emotion category between the low and high intensity expressions, separately for each expression and country. The expected frequencies were those calculated on the basis of the four (expression intensity level) \times nine (emotion categories) overall chi-squares separately for each emotion, poser, and country. Twenty-nine of the 32 comparisons were in the hypothesised direction, 20 of which were statistically significant. We interpreted these findings as providing strong support for Hypothesis 2.

To test Hypothesis 3—that Japanese will have higher agreement levels than Americans for low intensity expressions—we computed chi-squares on the target emotion category, using country as the independent variable, separately for each emotion and poser. For the low intensity expressions, only three of the 16 comparisons were significant—anger poser 4, sadness poser 1, and surprise

⁷ There has been some controversy in the literature with regard to the most appropriate null model to use in analysing categorical judgement data. Russell (1994), for example, has suggested that emotion judgements may occur according to dimensions such as valence and arousal, rather than through the use of emotion categories. Using this logic, he suggests that emotion terms may be grouped according to valence and arousal characteristics, thus altering the number of choices observers really have to choose from, thereby influencing the computation of chance. Other writers (e.g., Ekman, 1994; Izard, 1994), however, have argued against Russell's position, suggesting that judgement studies involving modifications to forced-choice categories produce little or no differences in findings. Some of our own previous research, involving categorical judgements but incorporating corrections for category usage (Biehl et al., 1997), and scalar ratings on multiple emotion categories (e.g., Biehl et al., 1997; Matsumoto & Ekman, 1989; Yrizarry, Matsumoto, & Wilson-Cohn, 1998) that allow for finer discriminations among expressions and examination of the clustering of emotion categories supposedly similar on valence or arousal, has suggested that judgements tend to provide no support for the Russell contention. Moreover, some of the impetus for the Russell position is mitigated in this study by the fact that the neutral and other categories are provided as response alternatives; his position directly attacked judgement procedures that did not allow for these possibilities. The Russell critique is also mitigated in this paper by the fact that expression intensity differences are tested on the same emotion category in the text immediately below. Nevertheless, this study was not designed to address these issues; rather, it addresses questions about judgments in the Ekman tradition, and readers should interpret the findings reported here within this line of work, and the caveats afforded by the controversy surrounding it in the literature.

TABLE 1

Percentage of observers selecting the emotion category intended by each expression

<i>Emotion</i>	<i>Poser</i>	<i>Country</i>	<i>Neutral</i>	<i>Low</i>	<i>High</i>	<i>Very high</i>	
Anger	1	USA	7.41	80.00	74.42	97.14	
	2	USA	11.11	62.96	80.65	76.74	
	3	USA	3.33	28.57	67.65	88.89	
	4	USA	0.00	41.67	48.15	66.67	
		Total USA	5.46	53.30	67.72	82.36	
	1	Japan	5.88	79.41	91.67	90.91	
	2	Japan	0.00	79.41	61.76	88.89	
	3	Japan	2.94	25.00	36.36	61.76	
	4	Japan	0.00	9.09	23.53	8.82	
		Total Japan	2.21	48.23	53.33	62.60	
		Total Anger	3.83	50.76	60.52	72.48	
	Happiness	1	USA	40.00	96.30	100.00	97.73
		2	USA	16.13	66.67	97.22	92.59
3		USA	55.56	83.33	97.73	97.22	
4		USA	6.82	91.67	96.30	100.00	
		Total USA	29.63	84.49	97.81	96.89	
1		Japan	42.42	85.29	97.06	100.00	
2		Japan	8.82	63.89	100.00	97.06	
3		Japan	23.53	85.29	83.33	87.88	
4		Japan	5.56	78.79	97.06	97.06	
		Total Japan	20.08	78.32	94.36	95.50	
		Total Happiness	24.86	81.41	96.09	96.19	
Sadness		1	USA	3.23	22.73	54.29	55.56
		2	USA	7.41	63.33	85.71	86.11
	3	USA	0.00	48.15	90.32	85.71	
	4	USA	2.33	33.33	74.07	90.32	
		Total USA	3.24	41.89	76.10	79.43	
	1	Japan	0.00	0.00	27.27	35.29	
	2	Japan	2.94	50.00	83.33	66.67	
	3	Japan	0.00	47.06	58.82	41.67	
	4	Japan	5.56	48.48	50.00	64.71	
		Total Japan	2.13	36.39	54.86	52.09	
		Total Sadness	2.68	39.14	65.48	65.76	
	Surprise	1	USA	3.70	68.97	95.35	97.22
		2	USA	0.00	70.37	96.55	93.18
3		USA	0.00	72.09	86.11	85.19	
4		USA	0.00	63.89	81.48	90.32	
		Total USA	0.93	68.83	89.87	91.48	
1		Japan	2.94	85.29	91.67	96.97	
2		Japan	0.00	20.59	97.06	97.22	
3		Japan	0.00	80.56	100.00	100.00	
4		Japan	2.78	93.94	100.00	100.00	
		Total Japan	1.43	70.10	97.18	98.55	
		Total Surprise	1.18	69.46	93.53	95.01	

poser 2; we thus rejected Hypothesis 3. The same tests for all other expressions produced the same findings; only 3 of 48 comparisons were statistically significant. Thus, the two countries did not differ in their categorical judgements, contrary to previous findings (e.g., Matsumoto, 1992).

Additional analyses. The intended emotion category was the modal category selected, and produced statistically significant chi-squares, for 62 of the 64 comparisons for high and very high intensity expressions, indicating that observers tended to choose the intended emotion category for them as well. Comparisons of the percentage of observers selecting the target emotion category between these expressions did not, however, suggest that the very high intensity expressions were associated with appreciably greater increases in agreement. Nine comparisons were statistically significant that indicated that the very high intensity expressions had higher agreement levels. But, four comparisons were significant and in the opposite direction. The remaining 19 comparisons were not significant.

Because the No Emotion and Other categories were unique to this study, we were interested in the degree to which they were actually used. The No Emotion category was used only 4% of the time (150 of 3360 total possible occurrences) when participants judged low, high, and very high expressions. Likewise, the Other category was used only 4% of the time (187 of 4384 possible occurrences). We concluded that the inclusion of these categories did not appreciably affect the findings.

Intensity ratings: Hypothesis 4

Overall analyses. A six-way ANOVA was computed on the intensity ratings, using country (2), gender (2), and expression intensity (4) as between subject variables, and emotion (4), poser (4), and rating type (2) as repeated measures. The country main effect was statistically significant, $F(1, 242) = 35.155$, $p < .001$, suggesting the possibility that cultural response sets existed.⁸

⁸The significant country main effect, which indicated that Americans had significantly higher ratings in general than the Japanese, was not the only evidence of the cultural response sets. The country by rating type interaction, which is the focus of these analyses, was also significant, $F(1, 242) = 25.78$, $p < .001$, indicating that Americans had higher mean ratings than the Japanese regardless of rating type. The significant country by rating type by intensity interaction, $F(3, 242) = 4.63$, $p < .05$, indicated that this effect differed across all four expression intensities by degree, not direction. The significant country by rating type by emotion interaction, $F(3, 726) = 15.27$, $p < .001$, also indicated that the effect differed across all four emotions only by degree, but not by direction. Further, we compared American and Japanese means separately for each poser, expression intensity, emotion, and rating type. Of the 128 total comparisons (4 posers \times 4 intensities \times 4 emotions \times 2 rating types), Americans had a higher mean rating than the Japanese 108 times, of which 62 were statistically significant. Cumulatively, these findings indicate that Americans had higher ratings than the Japanese regardless of emotion, poser, expression intensity, or rating type, and collectively provide strong evidence for the existence of a cultural response set operating in these data.

To eliminate this possibility, we standardised each observer's ratings on both scales across all items to his/her country mean computed by averaging across all ratings, including external and internal intensity⁹ (see Matsumoto, 1994, for a fuller discussion of the treatment of cultural response set-influenced data). We then re-ran the six-way ANOVA on the standardised data.¹⁰

Main analyses. To test Hypothesis 4—that Americans will give significantly higher ratings to external display relative to internal experience for low intensity expressions, while there would be no difference between the two ratings for the Japanese—we examined the significant three-way interaction involving country, expression intensity, and rating type, $F(3, 242) = 4.175, p < .01, R^2 = .05$, which suggested that the country differences in the ratings were mediated by expression intensity.¹¹ Thus, we tested the simple effects of ratings, separately for both countries and all four expression intensities, collapsing across the other factors; the error term was that used in testing the significant three-way interaction from the overall analysis (Table 2).¹² Americans rated external displays of high and very high intensity expressions significantly higher than internal experience, replicating the findings of Matsumoto et al. (1999; for high intensity expressions). The nonsignificant findings for American ratings of neutral and low intensity expressions indicated that their rating differences were specific to high intensity expressions. The Japanese rated internal experience significantly higher than external display for low intensity expressions; the

⁹ That is, a mean for each participant was computed across all 32 ratings made (16 expressions \times 2 intensity ratings), and the country mean and standard deviation of this aggregated average was then computed. Each rating was then standardised to this aggregated mean and standard deviation within each country.

¹⁰ Of the 32 effects testing gender, only one was significant—the 6-way interaction between emotion, poser, rating type, country, gender, and intensity. This interaction was associated with a very small effect size, $R^2 = .02$. Thus, we interpreted these findings as suggesting that gender did not play a role in influencing the ratings, and will not be discussed further.

¹¹ R^2 in both text and table refer to the amount of total variance accounted for by the particular effect referred to, and is computed $SS_{\text{num}}/SS_{\text{total}}$.

¹² An alternative approach here would be to investigate the simple effects of country, separately for each rating type. Although this approach may seem to be a more direct test of country differences, we opt for the analysis of the simple effects of ratings separately for each country for several reasons. First, this analysis allows for a direct comparison with the findings of Matsumoto et al. (1999), who reported simple effects of ratings in explaining their between-country effects. We contend that this analysis allows for a better glimpse as to why the country differences occurred in the first place. Second, both types of analyses (i.e., simple effects of country or ratings) are not orthogonal to each other; thus, statistically, the same effects are analysed given the same significant interaction from which it originates. Third, this approach allows for the sign tests reported immediately below, which are also congruent with the findings reported by Matsumoto et al. (1999), and transform the differences from one of group tendencies to individual behaviours. We believe that this individual-level view of the data is also important.

TABLE 2
Simple effects analyses of rating type, separately for both countries and all four expression intensities

Country	Expression intensity	External display	Attributed internal experience	F	p	R ² effect size	R ² effect size after covariates removed	Proportion of original effect size accounted for by IC and SD	Standardised coefficient for IC/IAI	Standardised coefficient for SDS
USA	Neutral	-1.84 (2.16)	-1.78 (2.08)	0.04	n.s.	.00				
	Low	-0.59 (1.36)	-.87 (1.41)	0.94	n.s.	.07				
	High	1.37 (1.20)	.55 (1.50)	8.18	<.01	.24	.02	.92	-.103	.141
	Very high	1.84 (1.26)	1.05 (1.58)	7.55	<.01	.21	.03	.84	-.147+	.081
Japan	Neutral	-2.86 (1.71)	-2.69 (1.89)	0.32	n.s.	.01				
	Low	-1.54 (1.58)	-.95 (1.73)	4.28	<.05	.19	.01	.93	-.191*	-.064
	High	1.64 (1.39)	1.56 (1.50)	0.07	n.s.	.00				
	Very high	2.63 (1.26)	2.21 (1.48)	2.22	n.s.	.09				

nonfindings on the other expressions indicated that these differences were specific to low intensity expressions, and replicate the findings of Matsumoto et al. (1999; for high intensity expressions).¹³ Thus, Hypothesis 4 regarding differences on the low intensity expressions was rejected; in fact, these findings are exactly opposite what was reported earlier by Matsumoto et al. (1999).

Because differences in the categorical judgements may have confounded the intensity ratings, we computed three-way ANOVAs using country and gender as between subject factors and rating type as a repeated measure, separately for each low, high, and very high intensity expression, including only those participants who chose the intended emotion category. Across the 48 analyses (16 expressions \times 3 intensity levels), nine country by rating type interactions were significant. Simple effects analyses of rating type essentially replicated the findings reported above. For the nonsignificant findings, the reduced sample sizes in some cells due to the inclusion criteria resulted in substantially reduced power. Thus, we tallied mean differences in ratings for each expression separately, focusing only on the three significant effects reported in Table 2. For low intensity expressions, Japanese participants had higher attributed internal experience ratings than external display on 11/16 comparisons. For both high and very high intensity expressions, Americans had higher external display means than internal experience 15/16 and 15/16 times, respectively. Thus, the results reported earlier were not confounded by differences in categorical emotion judgements.

To investigate the degree to which the mean differences were reflective of individual differences as well, we computed sign tests on the three significant findings. Sixty-one per cent of the Japanese judges viewing low intensity expressions gave a higher rating to internal experience; 27% rated display higher than experience; the rest were ties. These differences were significant, $z(137) = 5.09$, $p < .001$. Sixty-five per cent of the American judges viewing high and very high intensity expressions gave higher ratings to the external displays; 27% and 26%, respectively, gave higher ratings to internal experience; and the remainder were ties. These differences were also significant, $z(136) = 5.61$, $p < .001$; and $z(136) = 5.47$, $p < .001$, respectively. Thus, the findings reported above were well represented on the level of individuals as well.

Other analyses. A number of other statistically significant effects emerged (a full report may be obtained from the first author). We focused on effects

¹³ To further bolster the interpretation of the reliability of these findings, and the rationale for analysing simple effects of ratings, we analysed the 2-way country \times rating interactions separately for each level of expression intensity as well. This interaction was not significant for the neutral expressions, $F(1, 272) = 0.390$, n.s. It was, however, significant for all other expression intensities: $F(1, 271) = 41.478$, $p < .001$; $F(1, 271) = 19.579$, $p < .001$; and $F(1, 271) = 4.367$, $p < .05$, for weak, high, and very high expression intensities, respectively.

involving the country by rating type interaction, as it is most related to Hypothesis 4. The country by rating type interaction was significant, $F(1, 242) = 23.40$, $p < .001$, $R^2 = .10$, and a simple effects analysis of rating type indicated that Americans tended to give significantly higher ratings to external display than attributed internal experience, $t(242) = 3.42$, $p < .001$, while there was no difference in ratings for the Japanese, $t(242) = 0.49$, n.s. A simple effects analysis of country also indicated that Americans had significantly higher ratings on external display than did the Japanese, $t(242) = 1.70$, $p < .05$, while the Japanese had significantly higher ratings on attributed emotional experience than did the Americans, $t(242) = 2.20$, $p < .001$.¹⁴

The country by rating type by emotion interaction was also significant, $F(3, 726) = 13.76$, $p < .001$. We thus examined the simple effects of ratings separately for each emotion and country (Table 3). Americans gave significantly higher ratings to external display than to attributed internal experience on angry, sad, and surprise expressions; the Japanese had significantly higher ratings on

TABLE 3
Simple effects of rating type separately for each country and emotion

<i>Emotion</i>	<i>Country</i>	<i>External display</i>	<i>Attributed internal experience</i>	<i>F</i>	<i>p</i>
Anger	USA	0.76 (1.37)	-.02 (1.66)	30.95	.000
	Japan	0.35 (1.26)	.50 (1.53)	2.05	n.s.
Happiness	USA	0.32 (1.22)	.30 (1.16)	0.03	n.s.
	Japan	0.55 (1.22)	.46 (1.38)	0.85	n.s.
Sadness	USA	-0.45 (1.41)	-.85 (1.48)	15.75	.000
	Japan	-0.95 (1.42)	-.65 (1.44)	8.41	.004
Surprise	USA	0.21 (1.35)	-.46 (1.50)	43.10	.000
	Japan	-0.07 (1.40)	-.17 (1.52)	1.21	n.s.

¹⁴ We recognise that these two simple effects analyses are not orthogonal; but, we present both here for the comparison with the results of Matsumoto et al. (1999).

attributed internal experience than to external display on sad expressions. There were no other significant effects.

Unpackaging the cultural differences in intensity ratings with IC and SD: Hypothesis 5

Main analyses. In accordance with the procedures described earlier and reported by Matsumoto et al. (1998), we first computed an effect size for each of the simple effects of rating presented in Table 2 (column 7). We then recomputed each analysis that produced a significant effect using ICIAI and SDS scores as covariates, and recomputed the effect size of the rating effect (column 8).¹⁵ We then computed the proportion of the original effect size accounted for by the two cultural variables by taking a ratio of the difference between the old and new effect sizes to the original effect size (column 9). The two cultural variables accounted for nearly all of the original rating effect size in all three analyses — 92%, 84%, and 93% for American differences on high and very high intensity expressions, and Japanese differences on low intensity expressions, respectively.

To determine the relative contribution of the two cultural variables to the ratings, three multiple regressions were computed, each corresponding to the significant rating effect reported in Table 2 (columns 10 and 11, Table 2). In each analysis, the dependent variable was the difference between the external and internal intensity ratings; the independent variables were ICIAI and SDS scores. For Japanese ratings of low intensity expressions, the multiple *R* approached significance, $R = .204$, $F(2, 134) = 2.899$, $p < .06$. The regression coefficient for SDS was not significant; the coefficient for ICIAI was, suggesting that individual differences in IC contributed significantly to the difference between external and internal ratings for Japanese on low intensity expressions. The regression for American ratings of high intensity expressions

¹⁵ There are other options for conducting these analyses. One would involve the use of ICIAI and SDS scores as covariates in tests of the simple effects of country, separately for each rating type. Another would involve the regression of ICIAI and SDS on raw intensity ratings, the computation of residual scores based on the regression equations, and the testing of country differences on the residual scores. We opt for the procedure described in the text, however, for two reasons. First, the testing of the simple effects of rating type allows for a direct comparison with the results of Matsumoto et al. (1999), in which rating type differences were uncovered separately for each country, and served as a basis for explaining between-country differences. Second, restricting the analyses to within-country estimation of the influence of ICIAI and SDS on the ratings avoids the issue described in Footnote 3 and earlier in the introduction as to whether or not these individual-level scores capture the larger construct of culture, and as such whether or not they should or even can be used in between-country comparison tests. That is, limiting the analysis to examining the effects of ICIAI and SDS within each country allows for a direct testing of the most salient difference uncovered earlier by Matsumoto et al. (1999), and is not influenced by the confound of social comparisons described by Peng et al. (1997).

also approached significance, $R = .189$, $F(2, 133) = 2.455$, $p < .09$. Neither regression coefficient, however, was statistically significant. For American ratings of very high expressions, the multiple R was not significant, $R = .180$, $F(2, 133) = 2.217$, n.s. The regression coefficient for ICIAI, however, approached significance. Collectively, these findings suggest a stronger role for contributions of IC differences to the differences in emotion judgements reported earlier.

Other analyses. We examined country-level differences on the two individual-level cultural variables by computing two two-way ANOVAs on the ICIAI and SDS scores, using country and gender as the independent variables. For ICIAI, no effect was significant. Although one would expect American-Japanese cultural differences on individualism versus collectivism to exist, a recent review of research indicates that an overwhelming majority of studies that actually measure and test these differences on the individual level find no differences between the two countries on this dimension (Matsumoto, 1999). For SDS, the country main effect was significant, $F(1, 271) = 187.568$, $p < .0001$, indicating that the Japanese were more status differentiating than the Americans ($M_{\text{Japan}} = 1.87$, $SD_{\text{Japan}} = .62$; $M_{\text{USA}} = 0.87$, $SD_{\text{USA}} = 0.51$). Neither of the remaining two effects was significant.¹⁶

DISCUSSION

This study provides the first cross-cultural evidence that emotional facial expressions presented substantially weaker than normal are nevertheless judged reliably, but at lower agreement levels. These data complement the well-established basis for the universal recognition of these expressions, and extend the findings of Hess et al. (1997). Together, they raise interesting questions concerning the nature of the judgement process, with implications for future

¹⁶ Because the testing of between-country differences on these individual-level measures of culture-related variables was not a main goal of this study, we restrict discussion of these differences here. On one hand, the nondifference between Americans and Japanese on ICIAI would argue in favour of Peng et al.'s (1997) argument that social comparison processes may be at work when respondents complete the surveys, diluting possible between-country effects. On the other hand, research reviewed by Matsumoto (1999) examining individual-level IC differences between the US and Japan involves not only questionnaire studies, but also behavioural experiments, in which the social comparison argument may not be valid. In addition, why differences in the predicted direction on the other scale, SDS, would occur if social comparison processes were in effect further complicates the matter. We have no interpretation of differences on SDS and nondifferences on ICIAI here, but strongly suggest that future theoretical and empirical work be devoted to fleshing out the issues more carefully in the linkage between individual-level measurements of culture and the global constructs of culture typically used. In any case, these issues do not invalidate the within-country level analysis of the effects of ICIAI and SDS as covariates to the simple effects of rating type, as reported in this article.

areas of research. What are the biological substrates of emotion recognition? Is there a facial prototype recognition programme genetically encoded and biologically stored, as suggested by Ekman's neurocultural theory of emotion (1972)? Or, does emotion recognition develop through learned associations between expressions seen by infants and children and the labels adults provide for them? What are the minimal cues of reliable information? Will blends, partials, or profiles produce reliable and accurate judgements? What are the minimum standards for any expression to produce reliable judgements above chance levels? These questions need to be addressed in the future.

Although not a main goal of this study, the data reported in Table 1 also reflect interesting differences among the emotions. Even at low intensity, happiness and surprise are much more easily recognised than the two negative emotions. This is consistent with previous research, and suggests that negative emotions require greater signal clarity for optimal recognition. That more muscles are involved to produce the negative expressions, and that those muscles include more subtle actions than that more easily seen in happiness or surprise (e.g., the pursing of the lips in anger versus the raising of the lip corners in happiness or the dropping of the jaw in surprise) suggests that these two negative expressions are more complex and ambiguous, reducing signal clarity. This notion may have implications for adaptive functioning. As a signal source, negative emotions may require greater differentiation because it is important for decoders of those emotions to discern more precisely the nature of the specific emotion being expressed in order to aid in coping. Such would not necessarily be the case for happiness or surprise. Future research examining the behavioural response correlates of emotion recognition on the part of perceivers can help address this notion.

The lack of cultural differences in categorical judgements, despite the reporting of such differences in other studies involving Americans and Japanese, and despite our rationale concerning the possible higher accuracy of Japanese when judging ambiguous stimuli, is intriguing. One possible interpretation may be the diminishing differences between the two countries on cultural dimensions such as IC. As mentioned above, elsewhere we have reviewed 17 studies comparing the United States and Japan on the IC construct, and concluded that most recent research unequivocally rejects the notion that the United States is more individualistic than Japan, while Japan is more collectivistic than the United States (Matsumoto, 1999). In the present study, too, there were no country-level differences on individual-level IC. Diminishing differences on cultural dimensions such as IC on the country level may lead to the elimination of country level differences in emotion recognition as well, despite the fact that such differences may have existed in the past.

We do not, however, agree with this interpretation, for the simple reason that the multiple regression analyses on the intensity rating data implicated a role for individual-level IC to moderate the cultural differences found on these

ratings. Although not an impossibility, one would have to speculate quite dramatically that a single cultural dimension produces an effect in one context, but no effect in another, for the same psychological domain (i.e., emotion recognition). Given that the nonfinding of cultural differences on categorical judgments in this study is only a single finding in relation to multiple findings in multiple other studies, we suggest that the difference be re-tested in future research before speculations about the nature of changing differences be entertained.

That Americans rated external displays significantly higher than internal experience for high intensity expressions while there was no difference for the Japanese replicate exactly the findings of Matsumoto et al. (1999). For low intensity expressions, there was no difference between the two ratings for Americans, while Japanese rated internal experience higher than external display. These data suggest that when Japanese observe relatively weak expressions, they may infer that the poser is feeling emotions more strongly, compensating for the display rule that is presumed to be operating. When judging strong expressions, however, the Japanese interpret the strength of the expression as signaling that it must be an appropriate context to display emotions as such; thus, they interpret the poser's feelings as commensurate with the strength of the external display.

Americans operate on different presumptions. Knowing that they have a display rule to express emotions in a suggestive and even exaggerated manner, they presume less subjective experience when presented with high intensity expressions, reckoning that the poser probably does not feel the emotion as strongly as presented. With a weaker expression, Americans may interpret a context that does not require such exaggeration, thereby inferring that the poser is feeling emotion at a level commensurate with that expressed. These interpretations about the nature of the findings for Americans and Japanese for high and low intensity expressions, particularly with regard to the influence of display rules, are speculations, and should be treated as hypotheses in future research, including studies that actually measure display rules and their behavioral manifestations.

IC and SD accounted for practically all of the differences observed between the ratings for the Americans and Japanese. That is, differences between the ratings did not exist once the correlation between the differences and the individual-level cultural variables were accounted for. The coefficients for IC in the multiple regressions suggested that people who endorse collectivistic tendencies tend to infer more internal experience relative to external displays, while people with individualistic tendencies tend not to do so. The strength of the correlations, however, as well as of the overall *R*s was relatively low. Moreover, the relatively small coefficients for SD may reflect the fact that the experimental design was one in which SD was not fully relevant; the contribution of SD to judgement differences may be larger if status differences are manipulated as part

of the context within which judgements are obtained. Future research will need to tease out these effects more effectively.

There is an apparent contradiction in the findings that contributes to a dilemma for the interpretation of the results with regard to the psychological measures of culture. On one hand, IC accounts for the within-country rating differences; on the other hand, due to the absence of cross-country differences on IC, it is difficult to argue that IC as a global construct is somehow responsible for the observed pattern of findings. The missing piece to this puzzle is the answer to this question: does individual-level measurement of IC correspond to cultural level dynamics? If the answer is yes, then individual differences on IC can account for within-culture differences in ratings, but US-Japan differences in global, cultural level dynamics cannot. If the answer is no, then individual differences on IC can account for within-culture differences in ratings, but US-Japan differences in global, cultural level dynamics may or may not. The problem is that currently we do not know the precise answer to this question, and this study was not designed to address it. Substantial work will be required in the future to determine the degree to which individual-level measurement of IC or any other psychological culture variable corresponds to global, cultural level tendencies.

Still, given this state of affairs, the utilisation of individual-level measures of psychological culture represent the field's current best approximation of examining the contribution of cultural variables to psychological processes. The pattern of findings obtained suggests that irrespective of whether or not global, stereotypic assumptions about IC differences between the two countries are accurate, individual-level measures of psychological variables related to culture within the two countries account for the between-group differences observed and reported in this study, in the classical sense of covariates. Future studies examining other cultures, presumably those with national-level differences on IC, as well as other individual-level measures of culture, are needed to explore the boundaries of this interaction.

The inclusion of low intensity expressions produced differences in not only the categorical judgements, in terms of level of absolute agreement among observers, but also in the nature of the intensity ratings, and cultural differences in these. The very high intensity expressions, however, for all intents and purposes, did not produce any appreciably different results in either categorical judgement or intensity ratings between or within the cultures. This suggests that these types of judgements of facial emotions do not necessarily vary in a linear fashion according to the intensity of the expression judged, as suggested by Hess et al. (1997). Rather, they suggest the existence of certain thresholds of expressions that lead to different types of judgements, and that cultures differ in exactly what these thresholds may be. Future studies examining other expression intensities and other judgment tasks will elucidate the nature and boundaries of such a concept.

These findings have several implications for future research. Paramount among these is the need to incorporate context into studies examining emotion judgements. The very nature of IC and SD, for instance, which accounted for a large proportion of the observed differences, beg the question of context. Although this study presented faces in a laboratory situation, it may very well be that different contexts elicit different judgement patterns, especially concerning ratings of external displays and inferences about subjective experience. These differences may be associated with differential relative contributions of IC or SD (or other dimensions), depending on the nature of the context. Different contexts may elicit different judgements that access different culturally relevant cognitions. Presenting faces in ingroup versus outgroup contexts, for example, may access IC related cognitions more than SD; faces presented in high-same-low context, however, may access SD-related cognitions. At this time, however, we have no information about what these processes may be like, and this study suggests that such research be initiated.

This study was not conducted without limitation. The laboratory context utilised, the reliance on questionnaires to measure the two cultural dimensions, the relatively small number of posers within each emotion, and the comparability of the two samples are some of the concerns associated with this study. Nevertheless, the findings are unique, interesting, and provocative, and suggest several lines of inquiry that are substantially different than the usual judgement studies of the past.

Manuscript received 7 December 1999

Revised manuscript received 21 August 2001

REFERENCES

- Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, 21, 3–21.
- Bond, M.H., & Tedeschi, J.T. (2001). Polishing the jade: A modest proposal for improving the study of social psychology across cultures. In D. Matsumoto (Ed.), *Handbook of culture and psychology* (pp. 309–324). New York: Oxford University Press.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation*. Lincoln, NE: 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. (1978) *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53, 712–717.

- Ekman, P., Sorenson, E., R., & Friesen, W.V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*, 86–88.
- Hall, E.T. (1996). *The hidden dimension*. New York: Doubleday.
- Hess, U., Blairy, S., & Kleck, R.E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, *21*, 241–257.
- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.
- Hofstede, G. (1983). Dimensions of national cultures in fifty countries and three regions. In J. Deregowski, S. Dziurawiec, & R. Annis (Eds.), *Expiscations in cross-cultural psychology*. Lisse: Swets & Zeitlinger.
- Hui, H.C. (1984). *Individualism-collectivism: Theory, measurement, and its relation to reward allocation*. Unpublished doctoral dissertation, University of Illinois.
- Hui, H.C. (1988). Measurement of individualism-collectivism. *Journal of Research in Personality*, *22*, 17–36.
- Izard, C.E. (1971). *The face of emotion*. New York: Appleton Century Crofts.
- Izard, C. (1994). Innate and universal facial expressions: Evidence from developmental and cross-country research. *Psychological Bulletin*, *115*, 288–299.
- Kooken, K., Yrizarry, N., Consolacion, T., Vogt, A., & Matsumoto, D. (1997, April). *Individual level correlations between individualism-collectivism and status differentiation*. Paper presented in poster at the 1997 Annual Convention of the Western Psychological Association, Seattle, WA.
- Kooken, K., Yrizarry, N., & Matsumoto, D. (1996, April). *Status differentiation scale: A new scoring procedure*. Paper presented as a poster at the 1996 Annual Convention of the Western Psychological Association, San Jose, California.
- Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, *20*, 92–105.
- Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion*, *14*, 195–214.
- Matsumoto, D. (1991). Cultural influences on facial expression of emotion. *Southern Communication Journal*, *56*, 128–137.
- Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, *23*, 72–84.
- 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. (1994). *Cultural influences on research methods and statistics*. Pacific Grove, CA: Brooks Cole.
- Matsumoto, D. (1999). Culture and self: An empirical assessment of Markus and Kitayama's theory of independent self-construal. *Asian Journal of Social Psychology*, *2*, 289–310.
- Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion (JACFEE) and neural faces (JACNeuF)* [Slides]. Human Interaction Laboratory, University of California, San Francisco, 401 Parnassus Avenue, San Francisco, CA, 94143.
- Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, *13*, 143–157.
- Matsumoto, D., Kasri, F., & Kooken, K. (1999). American-Japanese cultural differences in judgments of expression intensity and subjective experience. *Cognition and Emotion*, *13*, 201–218.
- Matsumoto, D., Takeuchi, S., Andayani, S., Kouznetsova, N., & Krupp, D. (1998). The contribution of individualism vs. collectivism to cross-national differences in display rules. *Asian Journal of Social Psychology*, *1*, 147–165.
- Matsumoto, D., Weissman, M.D., Preston, K., Brown, B.R., & Kupperbusch, C. (1997). Context-specific measurement of individualism-collectivism on the individual level: The IC Interpersonal Assessment Inventory (ICIAI). *Journal of Cross-Cultural Psychology*, *28*, 743–767.

- Mulder, M., Veen, P., Hijzen, T., & Jansen, P. (1973). On power equalization: A behavioral example of power distance reduction. *Journal of Personality and Social Psychology*, 26, 151–158.
- Mulder, M., Veen, P., Rodenburg, C., Frenken, J., & Tielens, H. (1973). The power distance reduction hypothesis on a level of reality. *Journal of Experimental Social Psychology*, 9, 87–96.
- Pelto, P.J. (1968). The differences between “tight” and “loose” societies. *Trans-Action*, April, 37–40.
- Peng, K., Nisbett, R., & Wong, N.Y.C. (1997). Validity problems comparing values across cultures and possible solution. *Psychological Methods*, 2, 329–344.
- Poortinga, Y.H., Van de Vijver, F.J.R., Joe, R.C., & Van de Koppel, J.M.H. (1987). Peeling the onion called culture: A synopsis. In C. Kagitcibasi (Ed.), *Growth and progress in cross-cultural psychology* (pp. 22–34). Berwyn, PA: Swets North America.
- 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.
- Seki, K., Yrizarry, N., Kookan, K., & Matsumoto, D. (1995, March). *The status differentiation scale (SDS)*. Paper presented in poster at the 1995 Annual Convention of the Western Psychological Association, Los Angeles, California.
- Singelis, T.M., Triandis, H.C., Bhawuk, D.P.S., & Gelfand, M.J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research*, 29, 240–275.
- Triandis, H.C. (1994). *Culture and social behavior*. New York: McGraw-Hill.
- Triandis, H.C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.
- Triandis, H.C., Bontempo, R., Villareal, M.J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self in-group relationships. *Journal of Personality and Social Psychology*, 54, 323–338.
- Triandis, H.C., Leung, K., Villareal, M.J., & Clack, F.L. (1985). Allocentric versus idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in Personality*, 19, 395–415.
- Van de Vijver, F.J.R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.
- Wallbott, H., & Scherer, K.R. (1986). How universal and specific is emotional experience? Evidence from 27 countries on five continents. *Social Science Information*, 25, 763–796.
- Yamaguchi, S. (1994). Empirical evidence on collectivism among Japanese. In U. Kim, H.C. Triandis, C. Kagitcibasi, S.C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications*. (pp.329–350). Newbury Park, CA: Sage.
- Yrizarry, N., Matsumoto, D., & Wilson Cohn, C. (1998). American and Japanese multi-scalar intensity ratings of universal facial expressions of emotion. *Motivation and Emotion*, 22, 315–327.

Copyright of Cognition & Emotion is the property of Psychology Press (T&F) and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.