The study of emotion judgments across cultures forms a crucial basis for the universality of facial expressions of emotion (Ekman, 1972, 1973; Ekman & Friesen, 1971; Ekman, Sorenson, & Friesen, 1969; Izard, 1971). Studies examining categorical judgments of emotions are especially prevalent, involving judgments of posed and spontaneous faces, voice, and body postures presented in a range of modalities (e.g., slides, photographs, video, audiotape) to judges from many countries and cultures (Elfenbein & Ambady, 2002).

One type of emotion judgment that has received attention in the literature is intensity ratings. Unlike categorical judgments, which examine what emotions are perceived, intensity ratings ask how much emotion is present. The first study to document cultural differences in intensity ratings of facial expressions was Ekman et al.’s (1987), who showed the universal expressions of anger, disgust, fear, happiness, sadness, and surprise to judges in ten cultures, who rated the intensity of the expressions. Post-hoc analyses indicated that...
Asians perceived significantly lower intensity in the faces than non-Asians.

To examine these findings further, it was necessary to create a stimulus set of facial expressions expressed by people of visibly different cultures but that were equivalent in their physical signaling properties (i.e., they had to have the same muscles innervated at the same intensity levels). These characteristics were necessary because when testing for cultural differences in judgments, observers in different cultures need to view stimuli expressed by people of their own as well as other cultures in a balanced design to test whether or not the observed differences are limited to judgments of people of their own or other cultures. Because the expressions used in Ekman et al.’s (1987) study were all of Caucasian individuals, Asians knew immediately that the expressors were not of their own culture. Thus the Asians may have given lower ratings because of politeness in not wanting to attribute strong emotions to the faces they saw or because of ignorance about the expressions of people from obviously different cultures. Because most judgment studies until Ekman et al.’s (1987) were concerned with cultural similarities, this methodological requirement was not an issue. But once researchers turned their attention to explicating the nature of cultural differences, the balance between the perceived culture of the expressor and the judge’s culture became a crucial issue (Matsumoto, 2002a).

We decided to follow the Asian v. non-Asian cultural difference in intensity ratings by examining American v. Japanese comparisons. Given the methodological requirement of a stimulus set involving Asian faces, we created the Japanese and Caucasian Facial Expressions of Emotion (JACFEE; Matsumoto & Ekman, 1988), which includes 56 expressions, all by different individuals, of the seven universal emotions (the six above and contempt), half of which were expressed by Caucasians, the other half by Japanese (half male and half female within each). Most importantly, all expressions within each emotion were equivalent in the physical signaling properties according to FACS coding, ensuring that any judgment differences across expressors could not be confounded by differences in their expressions.

In the first study using the JACFEE, American and Japanese judges saw the faces twice, first rating the intensity of seven emotion categories on each expression, and second rating its global intensity (Matsumoto & Ekman, 1989; Yrizarry, Matsumoto, & Wilson-Cohn, 1998). The results on both ratings were the same: Americans gave significantly higher ratings than the Japanese on anger, happiness, sadness, and surprise, regardless of the culture of the expressor (contempt, disgust, and fear were not analyzed). These findings suggested that ignorance or politeness were not the reasons why Asians had significantly lower ratings than non-Asians. Instead, cultural differences were probably due to differences in cultural rules of decoding emotions.

These findings have been replicated (Matsumoto, 1990, 1991), and have also been obtained in studies involving judgments by members of other cultures including Hungary, Poland, Vietnam, and Sumatra (Biehl et al., 1997). They appear specific to intensity ratings of emotion, as other types of ratings do not produce the same findings (Matsumoto & Kudoh, 1993). Researchers have generally interpreted these findings as occurring because of cultural display rules. That is, because Japanese are known to suppress their display of emotions, this rule probably leads them to suppress the interpretations of the strength of the emotions of others.

Ratings of External Displays v. Internal Feelings

One limitation of the previous studies was that it was impossible to know whether the judges rated the intensity of the external display – the strength of the expression – or the internal feeling – the strength of the presumed subjective experience of the expressor. This distinction is important because it probably mirrors more closely the judgments people make in real life. This limitation was addressed in a study involving American and Japanese judges who saw the JACFEE and made three
judgments: a categorical emotion judgment, the intensity of the external display, and the intensity of the subjective experience of the expressor (Matsumoto, Kasri, & Koo.ken, 1999). Americans gave significantly higher ratings than Japanese to external displays, but the Japanese gave significantly higher ratings to internal feelings. Within-culture simple effects analyses indicated that the cultural differences occurred because Americans gave significantly higher ratings to external displays than internal experience, but that the Japanese did not differ in their two types of ratings. Once again, these findings were interpreted as occurring because of cultural display rules: because the Japanese are more likely to suppress their true feelings, they would be more likely to infer more subjective experience in others compared to Americans, who do not have the display rule to suppress their emotions as much. The results clearly demonstrated the meaningfulness of distinguishing between the two types of ratings.

In a subsequent study we examined if the cultural differences in intensity ratings might be specific to the intensity level of the expressions being judged. All of the JACFEE expressions are full-face, high intensity expressions, and exit interviews of the Japanese participants suggested that they did not differ in their external and internal ratings because they inferred that the emotions must have been justified if the expressions were large and uninhibited. Likewise, debriefing of the Americans suggested that they inferred less internal intensity precisely because they were associated with high intensity expressions, reasoning that everyone knows that people exaggerate their expressions relative to their feelings. Thus, the expressors were probably not feeling the emotions as strongly as they showed on their faces.

To examine these possibilities, it was necessary for us to create stimuli of different intensity levels but that included the same muscle movements. This was possible by utilizing computer morphing technologies. Using the JACFEE expressions and their corresponding neutral expression by the same individual as 100% and 0% anchors, we created low (50%) and very high (125%) intensity expressions for four emotions (anger, happiness, sadness, and surprise), and compared American and Japanese judgments of these (Matsumoto et al., 2002). Americans and Japanese did not differ in their ratings of neutral expressions; for high and very high intensity expressions, the Americans rated external display significantly higher than internal experience, while there was no difference for the Japanese, replicating the previous finding. For low intensity expressions, there were no differences between the two ratings types for Americans; the Japanese, however, gave significantly higher ratings to internal experience than to external display.

These results were once again interpreted as occurring because of cultural display rules. On one hand, when judging low intensity expressions, Americans did not differ in their ratings because they most likely inferred that the expressor was probably feeling the emotion at a level commensurate with the expression. When judging high and very high intensity expressions, however, the Americans probably judged that the expressers were exaggerating their expressions relative to their true feelings, commensurate with such display rules. On the other hand, when the Japanese judged low intensity expressions, they probably interpreted that the person was suppressing their emotions and really felt more emotion than they were showing. When judging high and very high intensity expressions, they probably felt that the person may have been justified in displaying such strong emotions.

The Need to Measure Display Rules

The studies reviewed above were all limited by the fact that none actually measured cultural display rules despite their importance in the interpretation of every cultural difference in judgment observed. Previous interpretations have essentially been speculations with no basis in empirical fact. The only way to be certain that the cultural differences in intensity ratings occurred because of cultural display rules is to actually obtain display rules.
and emotion judgments in the same study and to assess the effect of the display rules on the cultural differences in ratings either through a regression or covariance analysis. The studies below do so.

It makes theoretical sense to link display rules with intensity ratings. The process of enculturation is wrought with occurrences in which behaviors are socialized, reinforced, or extinguished. During this time individuals not only learn which behaviors are appropriate and not for themselves; they are also creating cultural filters with which to judge and evaluate the behaviors of others (Matsumoto & Juang, 2013). What is appropriate and not for oneself becomes one of the standards by which the behaviors of others are judged. While this makes conceptual sense across a wide range of psychological phenomena, within the realm of emotion the appropriateness of behaviors is represented by display rules and cultural filters are represented by judgments of the emotions of others.

The Contribution of Psychological Culture

At the same time, there is a need to unpack the psychological contents of culture and to examine the degree to which those contents affect emotion judgments. Cross-cultural theorists have made considerable strides in advancing knowledge about the psychological dimensions that underlie culture, including individualism v. collectivism, status and power differentiation, uncertainty avoidance, masculinity and femininity, tightness v. looseness, and contextualization (Hall, 1966, 1976; Hofstede, 1980, 1984; Pelto, 1968; Triandis, 1995; Triandis, Bontempo, Villareal, Asai, & Lucca, 1988). Individual level measures of some of these constructs have been developed and their inclusion in research is important for several reasons. They allow researchers to check assumptions about cultural differences underlying their samples, which often constitute individuals of different national or ethnic groups (Matsumoto & Yoo, 2006). They allow researchers to use the psychological level measures of culture as covariates that allow for a determination of the degree to which they account for cultural differences. To be sure, individual level measures of psychological culture are not culture as a macro-social concept, and the results of such endeavors need to be interpreted with that caveat. But they do provide researchers with the best estimation of how culture, as an individual and psychological construct, affects their data.

In fact Matsumoto et al. (2002) included individual level measures of individualism v. collectivism and status differentiation in their study, and regression and covariance analyses indicated that these variables accounted for almost all of the original variance associated with the cultural differences in intensity ratings. At the same time, cultural display rules were not measured so it was impossible to know the degree to which the mediational effects of psychological culture could be explained by differences in display rules. In Study 1, therefore, we included three measures of psychological culture along with a measure of cultural display rules to unpack the cultural differences in intensity ratings of facial expressions of emotion.

Overview of the Current Studies

In Study 1, American and Japanese participants viewed the same expressions used in Matsumoto et al. (2002) and made the same three judgments as in that study: a categorical emotion judgment, a rating of the intensity of the external display, and a rating of the intensity of the internal experience. The participants also completed a measure of cultural display rules and three measures of psychological culture. We hypothesized a significant culture by expression intensity by rating type interaction as was found in Matsumoto et al (2002). We further hypothesized that covariate analyses would indicate that cultural display rules and psychological culture would mediate the cultural differences in ratings.

In Study 2, we extended the findings from Study 1 by examining the degree to which a concept related to cultural display rules, emotion regulation, would mediate an expression level by rating type interaction in a sample of American judges.
Study 1

Methods

Participants. The participants were 151 Americans (101 females, 50 males, mean age = 24.86) and 78 Japanese (33 females, 45 males, mean age = 20.95). The participants were recruited from psychology classes at large universities in San Francisco and Tokyo, respectively. All participated voluntarily and were born and raised in their respective countries, with English and Japanese their primary languages.

Facial stimuli. The stimuli were 64 expressions adapted from Matsumoto and Ekman’s (1988) Japanese and Caucasian Facial Expressions of Emotion (JACFEE), which were used in Matsumoto et al. (2002). Four expressors were randomly selected from the angry, happy, sad and surprised expressions and matched with their corresponding neutral expression. Seven hundred thirty nine reference points were placed onto a digitized version of each 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%, 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 expressions. All stimuli were inspected by coders trained in FACS to insure that they portrayed expressions that were anatomically possible and involved the same action units as the original expressions. The final set included 64 expressions – 4 emotions x 4 expressors x 4 intensity levels.

Judgment tasks. Participants made three ratings for each expression. The first was a categorical judgment 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 expressor using two 9-point scales labeled None, 0, to A Lot, 8. The instructions for these ratings were the same as those used in Matsumoto et al. (2002) 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 the facial expression and (3) indicate the intensity level of what you think the person is actually feeling using this 0 to 8 scale. Please choose only the numbers used on this scale.”

Display rules. The Display Rule Assessment Inventory (DRAI), used in a previous cross-national study of display rules (Matsumoto, Takeuchi, Andayani, Kouznetsova, & Krupp, 1998), was used to assess display rules. Participants rated how they would express 14 emotions in four relationship settings (family, close friends, colleagues, and strangers) in two domains – what they should do and what they actually do. The 14 emotion terms included each of the seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, surprise) and a synonym (hostility, defiance, aversion, worry, joy, gloomy, and shock, respectively). Participants selected one of seven response alternatives, based on a theoretical range of possible behavioral responses (Ekman & Friesen, 1975):
1. Express the feeling as is with no inhibitions (Express).
2. Express the feeling but with less intensity than one’s true feelings (Deamplify).
3. Express the feeling but with more intensity than one’s true feelings (Amplify).
4. Remain neutral; express nothing (Neutralize).
5. Express the feeling but together with a smile to qualify one’s feelings (Qualify).
6. Smile only with no trace of anything else in order to hide one’s true feelings (Mask).
7. Some other response (Other).

Multidimensional scaling of the response alternatives by a separate group of participants indicated that a single dimension of “control” described the differences among the responses (Matsumoto et al., 1998). The nominal data were thus recoded into the scalar values associated with each of the responses on the dimension, with higher scores reflecting greater control exerted by an individual to achieve the expression. Scores were then averaged across both synonyms for each emotion (alphas across emotions within each social relationship ranged from .90 to .93; alpha for total score used in covariate analyses below .95).

**Psychological culture.** Participants completed the Individualism-Collectivism Interpersonal Assessment Inventory (ICIAI) (Matsumoto, Weissman, Preston, Brown, & Kupperbusch, 1997), the Status Differentiation Scale (SDS) (Matsumoto, 2007), and the Self-Construal Scale (SCS) (Singelis, 1994).

The ICIAI consists of 19 items rated in relation to four social relationships: family, close friends, colleagues, and strangers. Participants use a 7-point scale, with higher scores reflecting greater collectivism. Scores are computed by averaging across items within each relationship (alphas ranged from .81 to .89; alpha for total score used in covariate analyses below .92).

The SDS is a 20-item scale reflecting attitudinal and behavioral tendencies in relation to interactions with people at three different status levels (same, higher, lower) in two domains (work and school). Scores are computed by averaging across items within each status level and domain (alphas ranged from .85 to .89; alpha for total score used in covariate analyses below .92).

The SCS consists of 30 items measuring independent and interdependent self-construals (15 items each). Participants use a 7-point scale to rate each item and scores are generated by averaging across the 15 items for both self-construals (alphas .78 and .75 for independent and interdependent construals, respectively).

**Demographic questionnaire.** All participants completed a demographic questionnaire that asked about sex, age, place of birth and upbringing, length of living in their country, number of hours worked per week, marital status, religious background and affiliations, ethnicity, languages spoken, socioeconomic status, and educational level.

**Procedures.** Participants were tested in small groups and were given the DRAI and the demographic questionnaire prior to their coming to the laboratory. The judgment task and questionnaires were counterbalanced so that half of the participants in each culture did the judgment task first while half did the questionnaires first. For the judgment task, the task and answer sheets were explained and any questions answered. Three expressions not used in the main stimulus set were then shown one at a time, and participants rated these in order to familiarize themselves with the task. When all participants understood the judgment task, each of the 64 expressions was shown one at a time, in a random order, for approximately 30 s each. Judges made their three judgments of each expression on answer sheets provided. When the participants completed the questionnaires, they were randomly ordered in a packet for each participant.

**Main Results**

**Cultural differences in intensity ratings.** Table 1 shows the percent of observers selecting the emotion category intended by each expression. We summed the intensity ratings across all four expressors within each emotion, expression intensity, and rating type and computed a five-way Analysis of Variance (ANOVA) using country (2) and sex (2) as between subject factors and emotion (4), expression intensity (4), and rating type (2) as within subject factors. For the main analyses we were concerned only with interactions of country and rating type. As predicted, the expression intensity by country by rating type interaction was significant, $F(3, 477) = 28.78, p < .001, R^2 = .12$. The country by rating type interaction was also significant, $F(1,159) = 5.74, p < .01, R^2 = .03$. No other effects involving the country by rating type interaction were significant.
## Table 1
Percent of Observers Selecting the Emotion Category Intended by Each Expression

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Expressor</th>
<th>Judge Country</th>
<th>Neutral (0%)</th>
<th>Weak (50%)</th>
<th>Strong (100%)</th>
<th>Very Strong (125%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>LR</td>
<td>USA</td>
<td>0.7</td>
<td>68.7</td>
<td>82.1</td>
<td>89.3</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>USA</td>
<td>8.0</td>
<td>63.9</td>
<td>88.0</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td>BM</td>
<td>USA</td>
<td>0.7</td>
<td>40.0</td>
<td>63.5</td>
<td>65.6</td>
</tr>
<tr>
<td></td>
<td>AF</td>
<td>USA</td>
<td>4.6</td>
<td>45.0</td>
<td>74.8</td>
<td>78.2</td>
</tr>
<tr>
<td></td>
<td>Average USA</td>
<td></td>
<td>3.5</td>
<td>54.4</td>
<td>77.1</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>JPN</td>
<td>0.0</td>
<td>80.8</td>
<td>90.9</td>
<td>92.3</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>JPN</td>
<td>6.5</td>
<td>87.2</td>
<td>91.0</td>
<td>86.7</td>
</tr>
<tr>
<td></td>
<td>BM</td>
<td>JPN</td>
<td>2.6</td>
<td>16.9</td>
<td>18.2</td>
<td>35.9</td>
</tr>
<tr>
<td></td>
<td>AF</td>
<td>JPN</td>
<td>1.3</td>
<td>47.4</td>
<td>68.8</td>
<td>74.0</td>
</tr>
<tr>
<td></td>
<td>Average JPN</td>
<td></td>
<td>2.6</td>
<td>58.1</td>
<td>67.2</td>
<td>72.0</td>
</tr>
<tr>
<td></td>
<td>Average Anger</td>
<td></td>
<td>3.1</td>
<td>56.3</td>
<td>72.2</td>
<td>75.2</td>
</tr>
<tr>
<td>Happiness</td>
<td>EA</td>
<td>USA</td>
<td>47.9</td>
<td>87.4</td>
<td>95.4</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>TA</td>
<td>USA</td>
<td>62.4</td>
<td>94.6</td>
<td>97.3</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>JL</td>
<td>USA</td>
<td>11.3</td>
<td>92.1</td>
<td>97.3</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>LK</td>
<td>USA</td>
<td>9.3</td>
<td>90.7</td>
<td>96.7</td>
<td>97.3</td>
</tr>
<tr>
<td></td>
<td>Average USA</td>
<td></td>
<td>32.7</td>
<td>91.2</td>
<td>96.7</td>
<td>98.0</td>
</tr>
<tr>
<td></td>
<td>EA</td>
<td>JPN</td>
<td>39.7</td>
<td>64.1</td>
<td>89.7</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>TA</td>
<td>JPN</td>
<td>64.5</td>
<td>92.3</td>
<td>98.7</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>JL</td>
<td>JPN</td>
<td>18.2</td>
<td>75.3</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>LK</td>
<td>JPN</td>
<td>11.5</td>
<td>80.8</td>
<td>100.0</td>
<td>96.1</td>
</tr>
<tr>
<td></td>
<td>Average JPN</td>
<td></td>
<td>33.5</td>
<td>78.1</td>
<td>97.1</td>
<td>97.7</td>
</tr>
<tr>
<td></td>
<td>Average Happiness</td>
<td></td>
<td>33.1</td>
<td>84.7</td>
<td>96.9</td>
<td>97.9</td>
</tr>
<tr>
<td>Sadness</td>
<td>NH</td>
<td>USA</td>
<td>4.7</td>
<td>63.6</td>
<td>91.9</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td>CF</td>
<td>USA</td>
<td>5.3</td>
<td>65.3</td>
<td>78.7</td>
<td>81.5</td>
</tr>
<tr>
<td></td>
<td>JC</td>
<td>USA</td>
<td>3.3</td>
<td>34.4</td>
<td>68.9</td>
<td>80.1</td>
</tr>
<tr>
<td></td>
<td>DC</td>
<td>USA</td>
<td>2.0</td>
<td>25.2</td>
<td>67.3</td>
<td>68.2</td>
</tr>
<tr>
<td></td>
<td>Average USA</td>
<td></td>
<td>3.8</td>
<td>47.1</td>
<td>76.7</td>
<td>79.3</td>
</tr>
<tr>
<td></td>
<td>NH</td>
<td>JPN</td>
<td>2.6</td>
<td>46.2</td>
<td>78.2</td>
<td>65.4</td>
</tr>
<tr>
<td></td>
<td>CF</td>
<td>JPN</td>
<td>3.9</td>
<td>60.3</td>
<td>55.8</td>
<td>62.3</td>
</tr>
<tr>
<td></td>
<td>JC</td>
<td>JPN</td>
<td>7.8</td>
<td>51.9</td>
<td>41.0</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td>DC</td>
<td>JPN</td>
<td>0.0</td>
<td>11.7</td>
<td>57.1</td>
<td>46.1</td>
</tr>
<tr>
<td></td>
<td>Average JPN</td>
<td></td>
<td>3.6</td>
<td>42.5</td>
<td>58.0</td>
<td>58.2</td>
</tr>
<tr>
<td></td>
<td>Average Sadness</td>
<td></td>
<td>3.7</td>
<td>44.8</td>
<td>67.4</td>
<td>68.8</td>
</tr>
<tr>
<td>Surprise</td>
<td>AG</td>
<td>USA</td>
<td>1.4</td>
<td>56.7</td>
<td>85.2</td>
<td>82.1</td>
</tr>
<tr>
<td></td>
<td>JG</td>
<td>USA</td>
<td>0.0</td>
<td>32.4</td>
<td>97.4</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td>KK</td>
<td>USA</td>
<td>3.3</td>
<td>67.5</td>
<td>89.3</td>
<td>90.1</td>
</tr>
<tr>
<td></td>
<td>MM</td>
<td>USA</td>
<td>1.3</td>
<td>68.0</td>
<td>94.0</td>
<td>95.3</td>
</tr>
<tr>
<td></td>
<td>Average USA</td>
<td></td>
<td>1.5</td>
<td>56.2</td>
<td>91.5</td>
<td>90.2</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>JPN</td>
<td>2.6</td>
<td>43.6</td>
<td>83.1</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>JG</td>
<td>JPN</td>
<td>0.0</td>
<td>28.2</td>
<td>93.6</td>
<td>96.1</td>
</tr>
<tr>
<td></td>
<td>KK</td>
<td>JPN</td>
<td>3.8</td>
<td>88.3</td>
<td>97.4</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>MM</td>
<td>JPN</td>
<td>0.0</td>
<td>69.2</td>
<td>94.8</td>
<td>97.4</td>
</tr>
<tr>
<td></td>
<td>Average JPN</td>
<td></td>
<td>1.6</td>
<td>57.3</td>
<td>92.2</td>
<td>94.5</td>
</tr>
<tr>
<td></td>
<td>Average Surprise</td>
<td></td>
<td>1.6</td>
<td>56.8</td>
<td>91.9</td>
<td>92.4</td>
</tr>
</tbody>
</table>
We then computed four country by rating two-way ANOVAs separately for each expression intensity (Figure 1). All four were statistically significant or marginally significant, \(F(1, 211) = 5.51, p < .05, R^2 = .03; F(1, 213) = 3.18, p < .08, R^2 = .02; F(1, 213) = 14.76, p < .001, R^2 = .07; \) and \(F(1, 213) = 21.57, p < .001, R^2 = .09, \) for neutral, low, high, and very high intensity expressions respectively. Simple effects analyses indicated that on neutral expressions there were no differences between ratings of external display and internal feelings for Americans, \(F(1, 86) = 2.67, \) ns, \(R^2 = .03; \) the Japanese, however, rated internal feelings higher than external display, \(F(1, 77) = 19.44, p < .001, R^2 = .20. \) On low intensity expressions, both Americans and Japanese gave higher ratings to internal feelings compared to external displays, but the size of these differences were significantly larger for the Japanese, \(F(1, 136) = 9.25, p < .01, R^2 = .03; \) \(F(1, 136) = 15.66, p < .001, R^2 = .10 \) for Americans, \(F(1, 77) = 31.55, p < .001, R^2 = .20, \) and \(F(1, 77) = 54.42, p < .001, R^2 = .42 \) for Japanese, for high and very high intensity expressions respectively.

**Cultural differences in display rules.** We computed a five-way ANOVA on the DRAI using country (2) and gender (2) as between subject variables and domain (2), emotion (7), and relationship (4) as within subject variables. Three effects involving the country factor were significant. The domain by country interaction, \(F(1, 193) = 11.00, p < .001, R^2 = .05, \) indicated that Japanese rated more control than did Americans in relation to what they should do, \(F(1, 208) = 6.69, p < .05, R^2 = .03; \) but there was no difference in ratings of what they actually do, \(F(1, 203) = .04, \) ns. The emotion by country interaction, \(F(6, 1158) = 2.61, p < .05, R^2 = 0, \) led to the finding that the Japanese reported greater control for anger and contempt, \(F(1, 212) = 4.26, p < .05, R^2 = .02, \) and \(F(1, 207) = 6.84, p < .01, \)
The effects of display rules on cultural differences in intensity ratings. We then recomputed the overall five-way ANOVA on the intensity ratings using the total DRAI score as a covariate. We used the total DRAI score because all cultural differences found on the DRAI were always in the same direction (Japanese exerting more control than Americans) and because all items were highly correlated with each other (total DRAI alpha = .95). The country by rating by expression intensity interaction was still significant, $F(3, 618) = 8.83, p < .001, R^2 = .037$, but the effect was much smaller than the original analyses, indicating that display rules mediated the relationship between culture and rating differences (Baron & Kenny, 1986). Comparison of the original effect size to the recomputed effect size indicated that 69% of the original effect was explained by country differences on display rules. In particular, the rating by total DRAI interaction was statistically significant for both strong and very strong expressions, $F(1, 210) = 5.68, p < .05, R^2 = .026$; and $F(1, 210) = 4.59, p < .05, R^2 = .021$, respectively.

Cultural differences in psychological culture. We computed a three-way ANOVA on the ICIAI ratings using country (2) and gender (2) as between subject factors and relationship (4) as a within subject factor. The country main effect was significant, $F(1, 225) = 8.64, p < .01, R^2 = .04$, indicating that Americans were more collectivistic than Japanese. The country by relationship interaction was also significant, $F(3, 675) = 25.20, p < .001, R^2 = .11$; Americans endorsed collectivistic values more than did the Japanese in relation to family and close friends, $F(1, 227) = 48.23, p < .001, R^2 = .18$, and $F(1, 227) = 24.74, p < .001, R^2 = .10$, respectively.

We computed a three-way ANOVA on the SCS ratings using country and gender as between subject factors and scale (independent v. interdependent) as a within subject factor. The country by scale interaction was significant, $F(1, 210) = 7.60, p < .01, R^2 = .04$. Americans had significantly higher scores on the independent self-construal scale, $F(1, 215) = 25.79, p < .001, R^2 = .11$, but there was no difference on the interdependent scale, $F(1, 213) = 1.84, ns.$

We computed a four-way ANOVA on the SDS ratings using country and gender as between subject factors and setting (work v. school) and relationship (3) as within subject factors. The only significant effect involving country was the country main effect, $F(1, 219) = 68.17, p < .001, R^2 = .31$, indicating that Americans had significantly higher status differentiating scores than did the Japanese.

The contribution of psychological culture to the cultural differences in intensity ratings when display rules are incorporated. We computed total scores for the ICIAI and the SDS based on the significant main effects of culture on both scales and the fact that the intercorrelations among the items were high (alphas for total ICIAI = .92, for total SDS = .92). Pearson correlations between the total ICIAI, SDS, and both SCS scales with total DRAI separately for Americans and Japanese indicated that for Americans, SCS independent was negatively correlated with total DRAI, $r(143) = -.15, p < .05$, and total SDS was positively correlated with total DRAI, $r(148) = .14, p < .05$. No other correlations were significant.

We then recomputed the overall ANOVA on the intensity ratings, using total DRAI, total ICIAI, total SDS, and both SCS scores as covariates. The country by expression intensity by rating interaction was still significant, $F(3, 555) = 4.21, p < .01, R^2 = .02$, and comparison of the associated effect size with the original indicated that the psychological culture variables accounted for an additional 14% of the original culture by expression intensity by rating effect size over and above that accounted for by display rules.

Other judgment results. The rating type by expression intensity interaction, $F(3, 477) = 60.89, p < .001, R^2 = .24$, indicated that the cultural differences reported above qualified an effect that existed independent of culture. Simple effects analyses indicated that across both countries
judges tended to rate internal feeling higher than external display for neutral and weak expressions, but external displays higher than internal feelings for high and very high intensity expressions. When total DRAI was included as a covariate, the interaction was still significant, $F(3, 618) = 25.09, p < .001, R^2 = .10$, and effect size analyses indicated that display rules accounted for 58% of the variance in the interaction.

The emotion by country by rating type interaction was also significant, $F(3, 627) = 8.66, p < .001, R^2 = .03$. We thus computed four country by rating two-way ANOVAs separately for each emotion. The interactions were non-significant for anger and sadness, $F(1, 212) = 2.830$, ns, and $F(1, 213) = 1.565$, ns, respectively. For happiness and surprise, however, Japanese judges inferred significantly less intensity in subjective experience relative to external display compared to Americans, $F(1, 212) = 5.82, p < .05, R^2 = .03$, and $F(1, 212) = 9.36, p < .01, R^2 = .04$, respectively.

Because it was possible that judgments of non-intended emotions confounded the cultural differences in intensity ratings, we computed country by rating type ANOVAs separately for each expression including only those judges who selected the intended emotion category. Of the 48 comparisons ($4$ expressors x $4$ emotions x $3$ expression intensity levels; we chose not to analyze the neutral expressions because the associated percentages were negligible), $27$ were statistically significant and in the directions reported earlier. In addition computation of weighted means across the expressors and emotions within each expression intensity level indicated the same differences as those reported above. On low intensity expressions, there was no difference between American ratings of external display and internal feelings (means = $4.32$ and $4.46$, respectively); Japanese, however, had higher ratings of internal feelings (means = $3.18$ and $3.54$ for external and internal, respectively). On high and very high intensity expressions, both Americans and Japanese gave higher ratings to external display v. internal feelings, and the differences were larger for the Japanese (American means = $6.02$ and $5.87$ for high intensity expressions, and $6.49$ and $6.24$ for very high intensity expression; Japanese means = $5.38$ and $4.81$ for high intensity expressions, and $6.13$ and $5.28$ for very high intensity expressions).

**Discussion, Study 1**

That the Japanese inferred greater subjective experience relative to external display compared to American judges when judging low intensity expressions replicated the same finding reported by Matsumoto et al. (2002). The non-significant difference in rating type for American judges viewing neutral expressions also replicated the same finding reported previously. That the Japanese inferred greater subjective experience when judging neutral expressions is new to this study but consistent with the notion that Japanese will infer greater subjective experience to low intensity expressions. That Americans viewing high and very high intensity expressions rated external display higher than internal experience also replicated the same findings reported previously. That the Japanese also did so when judging high and very high intensity expressions is new to this study, and we have no post-hoc interpretations for it. That this discrepancy was significantly larger for the Japanese compared to Americans, however, nevertheless indicated a cultural difference on the rating types.

Covariance analyses indicated that display rules accounted for a substantial portion of the variance in the cultural differences in the ratings. This is, to our knowledge, the first evidence that directly links cultural differences in how people display emotions with how they perceive the emotions of others. The psychological culture variables accounted for an additional proportion of the variance in cultural differences in ratings above and beyond the effects of cultural display rules. These findings qualify those reported by Matsumoto et al. (2002) where psychological culture accounted for practically the entire effect of culture on ratings. This finding also suggests that there are additional aspects of psychological culture not related to display rules that contribute to the culture by rating type interaction, and that there are aspects of this interaction not related to psychological culture or
display rules at all. The finding also suggests the possibility that dimensions of psychological culture other than those measured in this study may account for differences in judgments. Future theorizing and empirical work would be necessary to flesh out what these additional influences may be.

Americans were more collectivistic than the Japanese on the ICIAI, were more status differentiating on the SDS, were no different than the Japanese on interdependent self-construals, and were no different in ratings of what they actually do with regard to their emotional displays. Although somewhat surprising, these findings are commensurate with other individual-level studies of cultural values and attitudes related to collectivism, individualism, self-concept, and interpersonal consciousness, including not only questionnaires in which social comparisons may occur but also experiments involving behavioral data, conducted by Japanese and non-Japanese researchers recruiting subjects from various areas of Japan all point to the same conclusion (Matsumoto, 2002b; Oyserman, Coon, & Kemmelmeier, 2002). Surveys conducted by the Japanese government on lifestyle, health, and work-related attitudes and values of working adults and people across a wide range of ages also point to the same conclusion (Matsumoto, 2002b). On the level of individuals, Japanese may not be more collectivistic than Americans, while such differences may exist on the cultural level.

**Overview of Study 2**

In recent years the concept of emotion regulation (ER) has gained widespread importance. Roughly defined as the ability to control, manage, and modify one’s emotional experiences and expressions, ER can be achieved by a variety of mechanisms. Gross, for example, suggests that individuals can alter their appraisals of situations so as to manage their interpretations of the types of events that bring about emotion (Reappraisal) (J. J. Gross, 1998, 1999a, 1999b, 2002; J. J. Gross & Levenson, 1993). Alternatively people can also suppress their emotional expressions and reactions (Suppression).

The concept of display rules is related to ER because display rules also have to do with the management and modification of emotional expressions and reactions. That is, display rules can be considered one component of ER. Conceptually, as emotion involves a package of events including cognitions, expressions, physiological reactions, and subjective experience, ER should refer to the degree to which each of these components are regulated. In Gross’ model, for example, Reappraisal may refer to one way in which cognitive regulation is achieved. Similarly, display rules should refer to the way in which expression regulation is achieved.

If display rules are conceptually linked to ER, and if display rules mediate the ways in which individuals perceive the emotions of others, then ER may also mediate those judgments. That is, people may differentiate the intensity of other’s external displays of emotions relative to presumed subjective experience on the basis of how they themselves regulate their expressions relative to their feelings. Individuals who are more likely to amplify their expressions relative to feelings may be more likely to perceive greater degrees of external displays relative to internal feelings. Conversely, people who are more likely to suppress their expressions relative to their feelings may be more likely to perceive greater degrees of internal feelings relative to external displays.

We tested this idea in an American sample in Study 2. We were limited to testing American participants because, to our knowledge, there is no measure of ER that has been validated for use in other cultures. And in any case the effect sizes related to the rating type by expression intensity interaction in Study 1, especially compared to the effect size associated with the country by expression intensity by rating type interaction, suggested that the rating differences that occurred as a function of expression intensity were sufficiently robust to be tested in a single sample.

In Study 2, participants viewed the same stimuli as in Study 1 and made the same judgments. This time, however, they also completed two measures of ER. We hypothesized that the rating type by ex-
pression intensity interaction would again be significant and that the measures of ER would mediate those differences.

Study 2

Methods

Participants. Eighty-six undergraduates (61 females, 25 males, mean age = 22.97, sd = 6.27) were recruited from psychology classes at two universities in the San Francisco bay area and participated in partial fulfillment of class requirements. Their ethnicities were as follows: 43.0% Caucasian, 5.8% African or African-Americans, 19.8% Asian or Asian American, 7.0% Filipino, 5.9% Hispanic or Latinos, 1.2% Indians, and 1.2% Middle Easterners; 14.0% of the participants chose two ethnicities and 2.4% of the participants were unclassifiable.

Facial stimuli and judgment tasks. The facial stimuli were the 48 neutral, low, and high intensity expressions used in Study 1. We dropped the very high intensity expressions because both in Study 1 and in Matsumoto (2002) they were not associated with any findings that were different than what was found for high intensity expressions. The judgment tasks were the same three tasks utilized in Study 1 as well as previous studies (Matsumoto et al., 2002; Matsumoto, Kasri et al., 1999).

Emotion regulation measures. We used two measures of emotion regulation: the Emotion Expressivity Scale (EES) and the Emotion Regulation Questionnaire (ERQ). The EES is a 17-item scale that assesses individual differences in the extent to which people outwardly display their emotions (Kring, Smith, & Neale, 1994). Previous research has demonstrated its temporal and internal reliability, and convergent and discriminant validity. Participants respond using a 6-point Likert scale from “never true” to “always true.” After reverse coding negatively loading items, a total score is computed by averaging all items (alpha = .92); higher scores reflect greater expressivity.

The ERQ is a 10-item scale that assesses the extent to which individuals typically try to inhibit their emotion-expressive behavior (J. J. Gross & John, in press). Previous research has demonstrated its reliability and validity. It is composed of two subscales: Reappraisal and Suppression. Participants rate each item using a 7-point scale ranging from “strongly disagree” to “strongly agree.” The six reappraisal items and four suppression items are averaged to create a score for each (alphas = .77 and .82, respectively), with higher scores indicating greater reappraisal and suppression.

Procedures. The experiments were conducted in groups ranging in size from 10 to 22 people. Half the participants completed the EES, ERQ, and demographic information before the judgment task; the other half completed judgment task first. The instructions for the judgment task were exactly the same as those used in Study 1. As in Study 1 three expressions not included in the study were used as practice so that participants would understand the judgments. There were no problems understanding the nature of the judgment task.

Main Results

Replication of the rating differences. We computed a four-way ANOVA on the intensity ratings using gender (2) as a between subject variable and expression intensity (3), emotion (4), and rating type (2) as within subject variables. As predicted the rating type by expression intensity interaction was highly significant, $F(2, 164) = 21.42, p < .001, R^2 = .21$. Simple effects of rating indicated that there were no differences between external and internal ratings for neutral and low intensity expressions, $F(1, 85) = 2.75, ns, R^2 = .03$, and $F(1, 85) = 1.33, ns, R^2 = .01$, respectively. Judges did, however, rate external displays significantly higher than internal feelings when judging high intensity expressions (External mean = 6.35, sd = .89, internal mean = 5.82, sd = 1.03), $F(1, 85) = 28.93, p < .001, R^2 = .25$. The rating type by emotion by expression intensity interaction was also significant, $F(6, 492) = 9.09, p < .001, R^2 = .10$; follow-up analyses indicated, however, the same pattern of findings as indicated by the rating by expression intensity interaction with only differences in degree across the emotions.
The effects of ER on the rating differences.
We then recomputed the overall ANOVA using ERQ Suppression, ERQ Reappraisal, and the EES as covariates. This time the rating type by expression intensity interaction was not significant, \( F(2, 158) = 1.64, \text{ ns}, R^2 = .00 \), indicating that the emotion regulation measures strongly mediated the differences between the rating types across expression intensities (Baron & Kenny, 1986), accounting for 100% of the original rating type by expression intensity effect size. The rating type by emotion by expression intensity was also not significant, \( F(6, 492) = 1.73, \text{ ns}, R^2 = .02 \), indicating that individual differences in emotion regulation accounted for 80% of the original variance in these differences.

Other results. We examined the degree to which demographic variables may have influenced the findings in two ways. For all scalar data, Pearson correlations were computed between the demographic variable and the difference score between ratings of external display and internal feelings on the high intensity expressions. For nominal demographic data, we computed ANOVAs on the difference scores using the demographic categories as independent variables provided that there were \( n > 20 \) cases in each cell. None of the analyses from either method was statistically significant, indicating that demographic differences did not confound the findings reported above.

The rating type by emotion interaction was also significant, \( F(3, 246) = 5.04, p < .01, R^2 = .055 \). Follow-up analyses indicated that judges rated external displays higher than internal experience on anger, happiness, and surprise, \( F(1, 85) = 6.43, p < .05, R^2 = .07; F(1,85) = 4.87, p < .05, R^2 = .05; \) and \( F(1, 85) = 6.52, p < .01, R^2 = .07 \), respectively. When the emotion regulation scores were used as covariates, however, the interaction was not significant, \( F(3, 237) = .43, \text{ ns}, R^2 = .005 \), indicating that 91% of the original variance in this interaction was accounted for by ER.

Discussion, Study 2

The rating differences occurred in this study exactly as they did in Study 1 and in Matsumoto et al. (1999; 2002). On neutral and low intensity expressions, there were no differences between ratings of external display and internal experience. On high intensity expressions, however, judges rated external displays significantly higher than internal experience. Moreover, this effect was mediated strongly by individual differences in ER; when ER measures were included as covariates, the effect disappeared entirely. ER also mediated a number of other rating differences in the judgments.

Study 2 extends the findings of Study 1 by utilizing two measures of ER as a substitute for display rules, and supports the notion that the degrees to which individuals regulate their emotions in general, and their emotional expressions in particular, affect how they judge emotions in others. An argument could be made that the two rating types – of external displays and internal experience – are themselves judgments of others’ ER or display rules. That is, if ER and display rules refer to the degree to which people modulate their expressions relative to their experiences, the difference between the strength of one’s external display and true internal experience is reflective of one’s ER or display rules. Judgments of these can then be considered judgments of other people’s ER, and the findings of these studies therefore suggest that individual’s own, personal rules of display or emotion regulation influence how they perceive other people’s attempts at ER or display management.

General Discussion

This article reports the first evidence of a link between display rules and emotion regulation with differences in ratings of the external display v. presumed subjective experience of others. These studies were not conducted without limitation, including the limited number of emotions tested, the artificial nature of the stimuli involved, and the lack of context information about the expressions that may affect such judgments. Future studies should take care to extend these findings by testing other emotions, using videotapes of real-life expressions imbedded within context for a greater understanding of the na-
ture of the differences in intensity ratings and the mediational effects of display rules on them.

Regardless of these limitations, however, the judgment differences reported replicate previous findings and are reliable. That a link between display rules and judgments exists makes conceptual sense, as behaviors that are socialized during enculturation may serve as the basis for cultural and personal filters that are used to interpret and evaluate the behaviors of others in our world. This notion is congruent with studies of emotional intelligence that have demonstrated correlations between emotion identification and emotion management, two accepted components of emotional intelligence (Roberts, Zeidner, & Matthews, 2001). This notion is also congruent with studies examining the relationship between personality and emotion recognition accuracy (ERA) that have reported a correlation between the dimensions of openness and conscientiousness and ERA (Matsumoto et al., 2000). Assuming the dimensions of openness and conscientiousness themselves are related to meaningful real-life behaviors, such data also suggest a linkage between behaviors and judgments of others. And a link between display rules and judgments is congruent with findings from research dating 30 years earlier demonstrating a correlation between ERA and emotion expressivity (Lanzetta & Kleck, 1970; Levy, 1964; Zuckerman, Hall, DeFrank, & Rosenthal, 1976; Zuckerman, Larrance, Hall, DeFrank, & Rosenthal, 1979). To be sure, though, emotional intelligence and ERA focus on emotion categorical emotion judgments, not judgments of expression management as we obtained in the studies reported here. In that sense, therefore, the findings reported here are unique to the field.

These findings open the door to confirming these notions across the age span. Developmental research on display rules has been conducted for over 20 years (Gnepp & Hess, 1986; Saarni, 1979; Zeman & Garber, 1996), as well as developmental research on judgments of emotion (Ekman & Friesen, 1971; D. Gross & Harris, 1988; Izard, 1992; Matsumoto & Kishimoto, 1983). The findings from the studies reported here suggest a merging of these two lines of inquiry in the same participants across the age ranges to examine the degree to which display rules and emotion regulation are indeed linked to emotion judgments in socialization. Such links across the age range in socialization and enculturation, if they are to be found, would provide new platforms by which concepts such as ethnocentrism could be understood.

These findings also open the door to examinations of other similar links with other types of social and person perception. Social judgments, attributions, interpersonal perception, and other such processes are well studied topics in psychology, and future research should examine the degree to which one’s own behaviors mediate judgments of the same behaviors in others.

References


Πολιτισμός, Κανόνες Έκφρασης Συναισθήματος και Κρίσεις Συναισθήματος

DAVID MATSUMOTO¹, JUNGWOOK CHOI², SATOKO HIRAYAMA³

AKIHIRO DOMAE⁴ & SUSUMU YAMAGUCHI⁵

ΠΕΡΙΛΗΨΗ

Στο άρθρο αυτό περιγράφονται δύο μελέτες που δείχνουν ότι οι πολιτισμικοί κανόνες έκφρασης συναισθήματος (Μελέτη 1) και η συναισθηματική ρύθμιση (Μελέτη 2) συνδέονται με τις κρίσεις διαχείρισης των συναισθηματικών εκφράσεων των άλλων. Στη Μελέτη 1, Αμερικανοί και Ιάπωνες κριτές είδαν πρόσωπα σε τέσσερα επίπεδα έντασης, βαθμολόγησαν την ένταση της εξωτερικής έκφρασης και υπέθεσαν την εσωτερική εμπειρία. Επίσης, συμπλήρωσαν ψυχομετρικές κλίμακες πολιτισμικών κανόνων έκφρασης και ψυχολογικού πολιτισμού. Οι πολιτισμικοί κανόνες έκφρασης αποτέλεσαν το 69% της διακύμανσης στις πολιτισμικές διαφορές στην αξιολόγηση της έντασης έκφρασης του συναισθήματος. Ο ψυχολογικός πολιτισμός αποτέλεσε το 14% της διακύμανσης. Στη Μελέτη 2, οι Αμερικανοί κριτές είδαν τα ίδια πρόσωπα και έκαναν τις ίδιες βαθμολογήσεις. Αυτή τη φορά όμως, συμπλήρωσαν δύο κλίμακες της συναισθηματικής ρύθμισης. Η συναισθηματική ρύθμιση εξήγησε σχεδόν όλες τις διαφορές βαθμολόγησης σε όλα τα επίπεδα έντασης της έκφρασης. Οι μελέτες αυτές καταγράφουν την πρώτη τεκμηρίωση –εξ όσων μπορούμε να γνωρίζουμε– της σύνδεσης μεταξύ των κανόνων έκφρασης του ατόμου και τη συναισθηματική ρύθμιση και των κρίσεων διαχείρισης του συναισθήματος σε άλλους.

Λέξεις-κλειδιά: Πολιτισμός, κανόνες συναισθηματικής έκφρασης, συναισθηματική ρύθμιση, κρίσεις συναισθήματος, εκφράσεις προσώπου.

1. Διεύθυνση: David Matsumoto, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132. Correspondence concerning this article can be sent to David Matsumoto, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA, 94132, telephone (415) 338-1114, fax (510) 217-9608, or e-mail dm@sfsu.edu.
2. Διεύθυνση: San Francisco State University.
3. Διεύθυνση: San Francisco State University.
4. Διεύθυνση: University of Tokyo.
5. Διεύθυνση: University of Tokyo.