Emotion
2007, Vol. 7, No. 4, 736–744

1528-3542/07/$12.00 DOI: 10.1037/1528-3542.7.4.736

Judgments of Emotion From Spontaneous Facial Expressions of New Guineans

Pamela J. Naab and James A. Russell
Boston College

The claim that specific discrete emotions can be universally recognized from human facial expressions is based mainly on the study of expressions that were posed. The current study ($N = 50$) examined recognition of emotion from 20 spontaneous expressions from Papua New Guinea photographed, coded, and labeled by P. Ekman (1980). For the 16 faces with a single predicted label, endorsement of that label ranged from 4.2% to 45.8% (mean 24.2%). For 4 faces with 2 predicted labels (blends), endorsement of one or the other ranged from 6.3% to 66.6% (mean 38.8%). Of the 24 labels Ekman predicted, 11 were endorsed at an above-chance level, and 13 were not. Spontaneous expressions do not achieve the level of recognition achieved by posed expressions.

Keywords: emotion perception, facial expression

The idea that your emotion can be read from the expression on your face has been a central assumption in important theories of emotion (e.g., Ekman, 1980; Izard, 1971; Tomkins, 1962). Well-known studies seem to show that observers, whatever their cultural background, recognize the emotion expressed by certain facial expressions (e.g., Ekman & Friesen, 1971; Ekman et al., 1987; Izard, 1971). Questions have been raised, however, about the methods used in these studies (Russell, 1994, 1995; see also Ekman, 1994; Izard, 1994). Perhaps no single methodological problem undermines the conclusion of universal recognition but, taken cumulatively, the problems call for new evidence.

One especially questionable feature of these often-cited studies is that the observers were shown facial expressions that had been posed. These posed expressions were created with the sole purpose of conveying a single specific emotion, typically with exaggerated individual features, with an absence of distracting or irrelevant features, and with a combination of features that rarely occurs in spontaneous expressions (Carroll & Russell, 1997; Scherer & Ellgring, 2007). The final set of photographs shown to observers had then been selected from all the poses photographed—presumably selected to be as recognizable as possible. There is no reason to believe that exaggerated posed faces so selected are representative of facial expressions encountered every day. Studies of selected posed expressions do not provide information on how often an emotion results in any change in the face nor on how recognizable the emotion is when facial changes do occur. Questions must therefore be raised about the extent to which observers recognize the emotion conveyed by facial expressions produced spontaneously. After all, theories that posit a link between emotions and faces are about naturally produced spontaneous facial expressions rather than posed ones.

We know of just six studies[1] that tested recognition of a range of emotions from spontaneous facial expressions (Hess & Blairy, 2001; Motley & Camden, 1988; Wagner, 1990; Wagner, Lewis, Ramsay, & Krediet, 1992; Wagner, MacDonald, & Manstead, 1986; Yik, Meng, & Russell, 1998). The broadest such study was a direct comparison of posed to spontaneous expressions shown via still photographs (Motley & Camden, 1988). They found that observers' recognition of emotion was significantly higher for posed (81.4% agreement with emotion the expresser posed) than for spontaneous expressions (26% agreement with emotion induced and which the expresser reported feeling). The spontaneous expression of happiness was recognized (61.3%) more often than chance, but expressions of other emotions were not (16% on average).

Wagner et al. (1986) asked observers to label dynamic spontaneous expressions (shown via video) of expressers responding to emotion-eliciting slides. Once again, happiness was recognized frequently (48.4%), but no other emotions came close to 50%. In addition to happiness, recognition was above chance for two other emotions: disgust (22.69%) and anger (12.67%). Recognition

Pamela J. Naab and James A. Russell, Department of Psychology, Boston College.

We thank Jose-Miguel Fernandez-Dols, Maya Tamir, and Ali Banuazizi for comments on a draft. We also thank Sherri Widen, Nicole Nelson, Jenny Shek, Gillian Cohen, Ryan Winton, Katie Clark, Kristen Hewett, and Heather Snead for their help with data collection, coding, and manuscript preparation. This study was supported by the National Science Foundation Grant 5000590.

Correspondence concerning this article should be addressed to James A. Russell, Department of Psychology, 301 McGuinn Hall, 140 Commonwealth Avenue, Boston College, Chestnut Hill, MA 02467. E-mail: james.russell@bc.edu

[1] Other studies on recognition of emotion from spontaneous facial expressions do exist but provide little information for present purposes. Most were limited, for example, to a single emotion (Camras et al., 2002), to a pleasant–unpleasant distinction (Buck, Savin, Miller, & Caul, 1972), to responses to a single eliciting situation (Scherer & Ceschi, 2000), or to a single type of facial expression (Keltner & Bonanno, 1997). Other studies provided observers with more than just the facial expression. For example, Cupchik and Poulos (1984) provided judges both with the expressers' facial expression and with the eliciting stimulus (a slide).

failed to differ from chance for two emotions—sadness (16.7%) and fear (10.3%)—as well as for a neutral expression (17.7%). Recognition fell significantly below chance for the emotion of surprise (11.5%). Wagner et al.'s (1986) point of comparison ("chance") was the base rate of use of the target label. When simple random choice among all labels served as the point of comparison, only happiness and disgust were recognized at levels above chance.

Two follow-up studies used a similar experimental procedure and obtained similar results. Wagner (1990) used a pretest to obtain emotion terms that expressers spontaneously used to describe their emotional reactions to the slides. In the main study, these terms were incorporated into the forced-choice rating format both for those viewing slides and for the observers in the recognition phase. The percentage of observers who accurately selected the correct emotion term ranged from 2.2% to 36.7%. Recognition for only amusement, puzzlement, and peacefulness was above chance. In a similar study, Wagner et al. (1992) examined results separately for men and women. With eight emotions, 32 assessments of accuracy resulted. Of these, only five were statistically significant, three of which involved the detection of amusement.

Yik et al. (1998) asked adults to label still photographs of Chinese babies' spontaneous expressions of sadness, anger, surprise, disgust, and fear. Observers recognized the happy faces (74%) but showed poor recognition for the remaining faces (23%), which did not differ significantly from chance (16.6%).

The five studies just described thus found frequent failures to recognize negative emotions from the faces of others. However, the results cannot identify whether the problem lies in the expresser or in the observer. That is, a failure to recognize someone's emotion could occur because the emotion was not shown on his or her face in the first place. Or, it could be that the emotion was shown on the face, but the observers failed to recognize the emotion shown.

The final study of spontaneous expression is therefore especially interesting. Hess and Blairy (2001) selected as stimuli only those spontaneous facial expressions that contained a "visible expressive display (as assessed by an experienced FACS [Facial Action Coding System; Ekman & Friesen, 1978] coder" (p. 133). Sixteen selected expressions were shown to observers via videotape. Observers identified happiness 87% of the time, and negative emotions, on average, 56.5% of the time: 45% for anger, 43% for disgust, and 75% for sadness.

In summary, six studies with different methods all found that happiness or amusement is frequently recognized from a spontaneous facial expression. For other emotions, however, these studies found lower recognition of the emotion from spontaneous expressions than had been found with posed ones. Indeed, the first five studies reviewed showed a striking contrast between results for spontaneous and posed expressions. But when Hess and Blairy (2001) selected as stimuli only cases with a visible facial expression, recognition was noticeably higher, although it was still lower than what is typically found (for similar observers) with posed expressions. Taken together, the six studies suggest that lower recognition of specific emotions from spontaneous expressions has two sources: fewer expressive faces are produced in the first place, and the expressive faces that are produced are less recognizable as to specific emotion than are posed faces. This conclusion of course remains tentative.

As important as these six studies are, much remains unknown. The six studies were all limited to a small number of emotions. Ekman (1992) listed interest and embarrassment as basic emotions, for example, but spontaneous expressions of these emotions were not included in the six studies. Ekman (1980) observed different facial expressions for the different positive states of happiness, interest, and relaxation (although Ekman, 1992, doubted whether different positive states have different facial expressions). Also, blends of emotions may be more common than single emotions, but spontaneous expressions of blends were not included in the six studies reported previously.

The current study offers additional data on the topic of recognition of discrete emotions from spontaneous expressions by taking advantage of a remarkable set of photographs taken, analyzed, and labeled by Ekman (1980) himself. These beautiful, clear photographs captured spontaneous facial expressions of the South Fore people of Papua New Guinea as they went about their daily lives. Isolated from Western culture, the New Guineans were not aware of the purpose of a camera:

> There was an enormous advantage to being with a people who were not camera-shy. They did not know what a camera did so they were not self-conscious about it, and much of their social life was outdoors and easily seen. (Ekman, 1980, p. 11)

Based both on his knowledge of the expresser's situation and on an analysis of the facial muscle movements visible in the photograph, Ekman (1980) labeled the emotion conveyed by each expression. Thus, as in Hess and Blairy's (2001) study, use of these photographs rules out the explanation that there was no expression on the face in the first place. Facial expressions for happiness, sadness, fear, disgust, surprise, anger, interest, and embarrassment were available. According to Ekman (1980), each of these is a basic emotion with a distinct facial expression. Ekman (1980) showed expressions for three additional nonemotional states: perplexed, hesitant, and relaxed.

The current study offered observers a greater than usual number of emotion labels to choose from (including labels for the basic emotions just listed plus perplexed, hesitant, and relaxed, but without including synonyms. Options included three positive terms (happy, interested, and relaxed). Ekman also coded some facial expressions as conveying two emotions simultaneously. We allowed for the possibility that participants might find two or more emotions in any one face by asking participants to label how intensely each emotion appeared in each face.

Another advantage of this stimulus set is that Ekman (1980) selected these photographs in such a way that they provide a clear prediction about the level of agreement to be expected:

> Since these pictures show universal facial expressions, the message conveyed by each face will usually be quite obvious. In the captions to the plates I have added brief explanations of exactly how these emotions are registered on these faces (or for that matter, any face). (p. 11)

Method

Participants

Fifty Boston College undergraduates, aged 18–25 years, participated in return for course credit.

Materials

Facial expressions were shown as still black-and-white 5 x 7-in. photographs of 20 spontaneous facial expressions of members of the South Fore of New Guinea from *The Face of Man* (Ekman, 1980).

Selection of photographs. We selected facial expressions in an effort to include as many different emotions as possible, with a better balance between positive and negative emotions than is usually achieved. Because we wanted to manipulate ascribed sex,[2] we excluded photographs of persons who were obviously male or female (Plates 4, 6, 11, 16, 24, 31, 41). Because we were looking for a clear view of the entire face, we also excluded faces that were partially obstructed (Plates 10, 13, 15, 32, 37C-D), that contained a large proportion of shadowing (Plates 9, 18, 43), that were blurred (Plate 28), that were photographed at a side profile (Plates 11, 15, 18, 26A-B, 27, 32, 33A-C, 33E-F, 34A-C, 35, 41, 42, 44), or that were very small (Plates 29, 31, 33A, 33D, 38, 39, 40). We did not want too many photographs of the same emotion; therefore after selecting eight plates for happiness, interest, and embarrassment, we eliminated six others (Plates 2, 5, 10, 26C-D, 37). Finally, we excluded faces that Ekman did not explicitly label (Plates 46, 52). The 20 faces that were included are described in Table 1.

Procedure

Each participant received a booklet with general instructions followed by 20 rating sheets. On each sheet was a single facial expression plus a questionnaire. There were four different orders of presentation of the faces (two random and two counterbalanced) to minimize order effects. For another purpose, for half the participants all faces in the booklet were said to be of women and for the other half of males.

Single best emotion. For each face, participants were asked to circle the one word that best described the emotional state of the person shown. There were 12 choices: *happiness, sadness, anger, fear, interest, relaxed, embarrassment, disgust, contempt, surprise, perplexed,* and *hesitant*. The set of emotion labels included each of the words Ekman used to describe the faces plus one additional word (*contempt*), which was added because Ekman and Friesen (1986) had proposed contempt as a basic emotion with a unique facial expression. A pilot test was administered to gather additional emotion words for these 20 faces through a free-response format; the labels so produced were roughly synonymous with the ones already on the list and, therefore, no additional emotion labels were added. Thus, the final list of 12 labels presented observers with a set of distinct choices, with no popular choice absent.

Intensity ratings. For each face, participants were also asked to indicate "the intensity to which the emotion is present in the face" for each of the same 12 emotions. Rating was on a 5-point Likert scale ranging from 0 (*not present*) to 4 (*maximum intensity*).

Results

Ekman's Predictions

No clear method or standard exists to establish what constitutes recognition. We therefore examine various possibilities.

Endorsement. The most common index of recognition is the percentage of participants who endorse the predicted label, which in this case was Ekman's label. Table 2 shows these percentages for the 16 photographs for which Ekman predicted a single label. Mean endorsement was 24.2%. For the six "basic" emotions included (happiness, sadness, anger, surprise, interest, and embarrassment), the mean endorsement was 24.3%. For the more traditional basic emotions included here and listed by Ekman and Friesen (1971; happiness, sadness, anger, and surprise), mean endorsement was 35.7%. For the three nonemotional states (hesitant, perplexed, and relaxed), the figure was 24.0%. Random selection among the 12 labels provided would yield 8.3% endorsement. In a one-tailed test of difference of proportions ($N = 50$), endorsement overall for Ekman's predicted single-emotion label (.242), for six basic emotions (.243), for the four traditional basic emotions (.357), and for the nonemotion states (.240) were each significantly higher than chance (.083) at the .01 level. In separate analyses for individual faces (again a one-tailed test of difference of proportions; $N = 50$), out of the 16 single-emotion expressions, the proportion endorsing Ekman's predicted label was above chance (.083) for nine faces at the .05 level, whereas for seven faces there was no significant difference.

Blended expressions have rarely been studied, and so no standard means of assessing recognition exists. For the four blended expressions, Table 2 shows the endorsement for each predicted label separately. On average, 38.8% of participants selected one or the other of the two predicted labels. One of the blends (Plate 9) achieved the clearest success for Ekman's predictions: 66.6% of participants selected either fear or surprise for the photograph Ekman labeled a fear–surprise blend. Random selection of one of two labels would yield 16.7%. In a one-tailed test of difference of proportions ($N = 50$), endorsement overall for one or another of Ekman's predicted labels for blended expressions (.388) was significantly higher than chance (.167) at the .01 level. In separate analyses for individual faces (again one-tailed test of difference of proportions; $N = 50$), out of the eight predicted labels across the four blended expressions, two of these labels were selected significantly more frequently than chance and six did not differ significantly from chance. A double asterisk in Table 2 identifies the labels that were endorsed at above-chance levels.

Modal responses. Another measure of recognition is whether the modal response corresponded to Ekman's predicted label. Matches are shown in Table 2. For the 16 single predicted labels, five modal responses matched Ekman's label. For the basic emotions (happiness, sadness, anger, surprise, interest, and embarrassment), four of the 12 modal responses matched Ekman's label. For the more traditional basic emotions (happiness, sadness, anger, and surprise), three of the six modal responses matched Ekman's label. For the nonemotional states, one out of the four modal responses matched Ekman's label.

With the four blended expressions, the two most frequently chosen labels matched the two predicted labels for one face, matched one of the predicted labels for two faces, and matched neither predicted label for one face.

[2] Another purpose of this study was to examine the effect of attributed sex on the emotion seen in a face, but this aspect of the study failed to produce significant differences.

Table 1
Twenty Spontaneous Facial Expressions of New Guineans

Plate no.	Age	Sex	Ekman's predicted label	Situational description
				Single-emotion expressions
8F	Adult	F	Happiness	"The woman…smiles at me" (p. 23).
37B	Adult	F	Happiness	"My approach is noticed, and a slight smile appears" (p. 67).
14	Child	M	Sadness	"Typically, the inner corners of the eyebrow are raised in sadness…with some people, however, not the brows but the skin above the brow moves, producing a characteristic wrinkle pattern, as shown by the boy in this photograph" (p. 30).
17	Adult	F	Anger	"Paying attention, conspicuously, to a woman usually was met with embarrassment, but on this occasion the woman staring out was angry about it, for reasons I don't know" (p. 34).
7	Adult	F	Surprise	"By luck I had the camera already at my eye when I surprised this woman" (p. 22).
8M	Adult	F	Surprise	"I again caused surprise, this time in the woman holding her lower leg" (p. 23).
22	Child	M	Interest	"Interest, an emotion we did not study but which we believe is also universal in its expression" (p. 40).
23	Baby	?	Interest	"Often when a person is interested and absorbed in what he is looking at, the lips may be parted, but the jaw does not drop open as far as in surprise" (p. 41).
26A	Child	F	Interest	"In a few seconds this sequence of pictures was taken of a young girl watching in fascinated interest as members of her hamlet talked with one of the scientists" (p. 44).
26B	Child	F	Interest	
36	Child	F	Embarrassed	"The tight-lipped embarrassed or self-conscious smile" (p. 65).
33G	Child	F	Embarrassed	"A better view of that tight-lipped embarrassed smile" (p. 60).
20	Adult	F	Perplexed	"…perplexed-like face" (p. 38).
8C	Adult	F	Perplexed	"The woman in the foreground holding a child shows the perplexed-like look" (p. 23).
3	Child	M	Hesitant	"This boy does not look as if he is enjoying himself. He seems hesitant, uncertain about what is coming next or whether he can or should be happy" (p. 17).
45	Child	F	Relaxed	"…relaxation. Here the positioning of the bodies of these two young women, their posture, hand positions, and physical contact, add to the mood of gentle warmth and calmness. They were unaware of the photographer" (p. 79).
				Blended expressions
12	Child	M	Fear-Disgust	"Three slightly fearful or worried children watch my coworker talk to adults from their hamlet. The child at the rear shows fear in the brows and the eyes, but the mouth might be showing a slight disdain or disgust" (p. 28).
9	Baby	?	Fear-Surprise	"A child afraid of me, consoled by his mother…. The brows are raised in both fear and surprise…the mouth shows surprise as well as fear" (p. 25).
19	Adult	F	Perplexed-Anger	"A detailed picture of the perplexing "perplexed' brow. This expression was very common. I saw it more than any other, yet I could never be certain about what it meant…It could be also that the person is mildly annoyed, or more than annoyed but controlling the expression so that it shows only in the brow" (p. 37).
25	Baby	?	Fear-Interest	"Apprehension blended with interest, but probably more of the fear message than the interest one. Mother is amused" (p. 43).

Note. From *The Face of Man* by Ekman (1980). In Plate 8, there are three women pictured: 8C is the woman closest to the camera, 8F is the woman furthest from the camera, and 8M is the woman in the middle of the other two. F = female; M = male; ? = sex unknown.

Intensity ratings. Another measure of recognition can be had from the intensity ratings. Table 2 shows the mean intensities for Ekman's predicted label. For 5 of the 16 faces with single predicted labels, the emotion with the highest mean intensity matched Ekman's predicted label.

We asked whether there was a significant difference between the intensity rating given Ekman's predicted label and that for a comparison label (also shown in Table 2), either the modal response or, in cases where the modal response matched Ekman's label, the second most frequently endorsed label. There are three possible outcomes for each face: (a) Ekman's predicted label was rated as significantly more intense than the comparison label, (b) the comparison label was rated as significantly more intense than Ekman's label, or (c) no significant difference was found between the two. Table 2 shows a t statistic for each comparison. A positive t indicates higher rated intensity for Ekman's predicted label than for the comparison label. A negative t statistic indicates higher rated intensity for the comparison label. Of the 20 comparisons, 2 t tests were significant and positive, 8 were significant and negative, and the remaining 12 were nonsignificant.

For the four blended expressions, we tested whether intensity ratings were higher for Ekman's two labels than the comparison label. For each face, when neither of Ekman's labels was the most intense, we compared them to the most intense label (this occurred for one face). When one of Ekman's labels was the most intense, we compared the predicted labels to the second most intense label (this occurred for two faces). When both of Ekman's labels were most intense, we compared them to the third most intense label (this occurred for one face). Table 2 gives t test results ($N = 50$) indicating that intensity ratings for both Ekman's two predicted

Table 2
Agreement of Participants (in Percentage of Endorsement and Mean Intensity) on Ekman's Predicted Label(s) Versus a Comparison Label for Each Facial Expression

Face	Ekman's predicted label				Comparison label[a]				
Plate no.	Label	% endorsement	Mean intensity[b]	CI	Label	% endorsement	Mean intensity[b,c]	CI	*t* statistic (*df* = 49)
				Single basic emotions					
37B	Happiness	33.3**	1.47	1.06–1.88	**Sadness**	45.8	2.06	1.64–2.48	−1.35
08F	**Happiness**[d]	33.3**	2.12	1.82–2.42	Interest	25.0	2.02	1.72–2.32	0.66
14	**Sadness**[d]	45.8**	2.41	2.03–2.79	Perplexed	35.4	2.49	2.10–2.88	0.51
17	Anger	20.8	1.84	1.49–2.19	**Sadness**	54.2	2.53	2.17–2.89	−2.96*
07	Surprise	35.4**	2.49	2.14–2.84	**Perplexed**	39.6	2.75	2.43–3.07	−1.19
08M	**Surprise**[d]	45.8**	2.67	2.36–2.98	Perplexed	18.8	2.31	1.97–2.65	1.67
22	Interest	6.3	1.43	1.14–1.72	**Contempt**	35.4	1.94	1.55–2.33	−2.28*
23	Interest	16.7**	1.88	1.54–2.22	**Sadness**	20.8	1.47	1.13–1.81	−1.65
26A	**Interest**[d]	33.3**	2.37	2.04–2.70	Perplexed	18.8	1.82	1.49–2.15	2.81*
26B	Interest	6.3	1.69	1.38–1.99	**Happiness**	33.3	2.08	1.73–2.43	−2.04*
36	Embarrassed	10.4	0.75	0.42–1.08	**Happiness**	33.3	1.92	1.61–2.23	−5.36*
33G	Embarrassed	4.2	0.82	0.52–1.12	**Happiness**	66.7	3.02	2.75–3.29*	−10.94
M		24.3	1.83			40.6	2.28		
				Single nonemotion states					
20	Perplexed	25.0**	2.18	1.84–2.52	**Disgust**	29.2	2.41	2.09–2.73	−0.72
8C	**Perplexed**[d]	22.9**	1.68	1.30–2.06	Interest	10.4	1.52	1.18–1.86	0.81
3	Hesitant	18.8	1.27	0.94–1.60	Interest	25.0	1.98	1.67–2.29	−3.45*
45	Relaxed	29.2**	1.58	1.17–1.99	**Sadness**	39.6	2.02	1.62–2.42	−1.21
M		24.0	1.68			2.92	2.02		
				Blended emotions[e]					
12	Fear	00.0	0.82	0.61–1.03	Perplexed	18.8	1.41	1.01–1.81	−2.76*
	Disgust[d]	25.0	1.75	1.37–2.13	—	—	—	—	0.98
9	**Fear**[d]	45.8**	2.43	2.04–2.82	Perplexed	10.4	1.94	1.57–2.31	1.66
	Surprise	20.8	2.22	1.93–2.51	—	—	—	—	1.76
19	Perplexed	22.9	2.08	1.67–2.49	Contempt	25.0	2.12	1.73–2.51	−0.01
	Anger[d]	34.5**	2.67	2.34–3.00	—	—	—	—	2.55*
25	Fear	4.2	1.27	0.75–1.79	**Anger**	60.4	2.73	2.39–3.07	−4.44*
	Interest	2.1	1.43	1.14–1.72	—	—	—	—	−5.96*
M		38.8	3.67			60.2	4.23		

Note. $N = 50$ for the intensity rankings, but because two participants failed to follow instructions, analysis of % endorsement for single emotion labels is based upon 48 participants. CI = confidence interval.
[a] Comparison label is the label, other than Ekman's predicted label or labels, with the highest endorsement. Modal response is shown in bold. [b] Mean intensity ratings were measured on a 0 to 4 scale. [c] The highest intensity response matched the modal response for 16 out of 20 faces, therefore the mean intensity for the modal response was given; differences between the intensity levels of the highest intensity and the modal response were negligible. [d] Indicates that the modal response matched predicted label. [e]Means for blends were calculated using the sum of the two labels for each face. * Indicates that *t* test found the intensity of the modal response (or if the modal response equals was the predicted label, the second most intense response) was significantly greater than the intensity of the predicted label at .05 level. ** Indicates Ekman's predicted label was significantly higher than chance (.083) in a difference of proportions test of single emotions at the .05 level.

labels were significantly above that of the comparison label for one face, one predicted label was above the comparison labels for two faces, and no label was above the comparison label for one face.

Consensus Scoring

Endorsement of Modal Responses. Another possible measure of recognition sets aside the predicted label and simply relies on the modal response within this sample—a procedure known as *consensus scoring* (Barchard & Russell, 2006). Modal responses are shown in Table 2 for each face. Overall, 42.2% of participants selected the modal response. If participants selected randomly among the 12 possible labels, the expected mean would be 8.3%. In a one-tailed test of difference of proportions ($N = 50$), endorsement for modal responses (.422) was significantly different from chance (.083) at the .001 level.

Intensity. The modal response also received the highest intensity rating in 14 of the 20 faces. In cases where the two differed, the difference was negligible. Overall, the mean intensity of the modal response was 2.1 out of 4.

Discussion

Our study examined only a small part of the large domain of facial expressions (Russell & Fernandez-Dols, 1997). Our stimulus faces were few in number and limited to still black-and-white photographs, all of people from one culture. Our observers were all college-educated adults of another culture who were shown the full

set of facial expressions and given a short list of emotions from which to choose. The faces had been preselected as ones that convey a specific emotion or specific blend of emotions. We had no information on reactions to faces not selected.

Despite these limitations, our study does provide valuable information for understanding communication via the face. Our study is one of the very few that have examined recognition of a range of specific emotions from spontaneous expressions. It is one of only two that presented observers with spontaneous faces preselected to contain visible facial expressions. It is the only study we know of that did so with spontaneous expressions coded by Ekman (1980) himself as to the specific emotion implied by each face. Ekman had selected these expressions after he had coded them for emotion based on visible facial movements and on his knowledge of the context of the expression. Indeed, Ekman's purpose was to illustrate how facial expressions reveal emotion in a way that can easily be recognized by everyone.

In the traditional approach, to know whether emotions are recognized from facial expressions, we must meet two requirements. The first is to establish the expresser's emotion. There is no gold standard, but there are various partial solutions to this requirement, including self-reports of the emotion the expresser feels, assessment of the eliciting context, and assessment of behavioral indices of the expresser's emotion, including facial behavior (Fernandez-Dols, Sanchez, Carrera, & Ruiz-Belda, 1997). For this requirement, our study had, to our knowledge, a unique advantage of examining spontaneous facial expressions selected and coded by Ekman (1980) on the basis of the expresser's eliciting context and the actual facial behavior.

The second requirement is to set a threshold for the amount of agreement among observers that constitutes recognition. There is again no gold standard, and it is difficult to know how high to set the bar—the higher the bar, the stricter the standard. Some writers have pointed to the near unanimous agreement seen in the well-known cross-cultural studies, and so one might ask whether agreement on a spontaneous expression meets that standard. Another possible standard is a majority, another is the mode, and yet another is any level above chance. Conclusions drawn from our results depend on where the bar is set.

If the question is whether recognition of emotion is the same from spontaneous as it is from posed expressions, then one must set the bar high. For one set of eight studies based on a similar method of still photographs with Western observers using a forced-choice response format (see review by Russell, 1994), the average amount of agreement was 84.4%. Our study joins the previous six studies on spontaneous expressions (Hess & Blairy, 2001; Motley & Camden, 1988; Wagner, 1990; Wagner et al., 1986, 1992; Yik et al., 1998) in showing that spontaneous expressions produce less agreement than this standard. Our findings raise, but do not answer, the question of why.

As to why, one possible explanation lies in the nature of a posed expression, which is created for the purpose of conveying a single emotion, with each feature exaggerated, with an absence of distracting or irrelevant features, and with a combination of features that might rarely occur spontaneously (Carroll & Russell, 1997; Scherer & Ellgring, 2007).

Another (not incompatible) possibility is that all our facial stimuli were produced by New Guineans but judged by Westerners. Elfenbein and Ambady (2002a, 2002b) showed that groups often have an in-group advantage when judging emotion from facial expressions, and so Bostonians could be expected to be better at reading facial expressions of other Bostonians than of New Guineans. The in-group advantage could be due to slight differences in appearances of facial expressions (nonverbal accents; Marsh, Elfenbein, & Ambady, 2003) or to the interaction between the encoding and decoding of facial stimuli— both of which are accounted for in dialect theory (Elfenbein & Ambady, 2003; Elfenbein, Beaupre, Levesque, & Hess, 2007; Marsh et al., 2003). Dialect theory hypothesizes that, just as dialects differ across groups speaking the same language, expression of emotion may differ across cultures, causing somewhat lower recognition by those from a different culture. Nevertheless, we doubt that dialect accounts fully for the large effects observed here. Some studies do not find an in-group advantage (Beaupre & Hess, 2005), and some studies have even found out-group advantages for which recognition is higher for groups from a culture outside of the expresser's (usually found with those of minority status better recognizing the expressions of those of majority status; Elfenbein & Ambady, 2002b). Further, cultural differences in dialects are unlikely to have played a role in the previous studies on spontaneous expressions (Hess & Blairy, 2001; Motley & Camden, 1988; Wagner, 1990; Wagner et al., 1986, 1992; Yik et al., 1998). Nevertheless, the role of cultural advantages, accents, and dialects remains an intriguing empirical question. How fascinating it would be to have spontaneous expressions from two culturally different societies being judged by members of both societies.

In any case, whatever the explanation, the empirical result that spontaneous expressions yield lower amounts of agreement on specific emotions than do posed ones is highly relevant to psychological theories on facial expression. This general conclusion must be followed by such questions as whether recognition from spontaneous expression varies with emotion. For this specific question, rather than setting an absolute bar, we can compare relative results across emotions. Although we had too few representatives of each emotion to draw firm conclusions, our results are consistent with Haidt and Keltner's (1999) notion of a gradient of recognition. In their cross-cultural research, they found large differences in recognition for different emotions—a finding that echoes results with preschool children (Widen & Russell, in press). At one extreme in the current study, Plate 9, which was labeled by Ekman as a surprise–fear blend, was labeled either fear or surprise by 66.6%. At the other extreme, Plate 33G, which was labeled by Ekman as an expression of embarrassment, was labeled as embarrassment by 4.2% of observers.

When the question is simply whether our observers recognized the emotion in the faces we showed them, it is unclear where to set the bar. For the 16 faces for which Ekman predicted a single label, endorsement of that label ranged from 4.2% to 45.8%, with a mean of 24.2%. This result is far below the 84.4% found with posed expressions and below the 50% level. Perhaps the most disturbing finding of this study for advocates of the theory that facial expressions convey specific emotions came from consensus scoring. A majority of observers rarely agreed on any one specific emotion for a spontaneous face. For only 3 of 20 faces did a majority agree on a label: 54.2% labeled Ekman's anger face as sad, 66.7% labeled his embarrassed face as happy, and 60.4% labeled his fear/interest blend as angry. For the remaining 17 of the 20 faces, agreement fell below 50%, with observers choosing a variety of

labels. Thus, no matter what the "real" emotion was behind the face, a majority of observers did not recognize that emotion from a spontaneous expression.

Pondering the lack of consensus found here, one must ask if these results are informative about the reading of emotions via facial expressions in the world outside the laboratory. Of course, there are a number of reasons to suggest that recognition might generally be higher outside the laboratory where observers have a variety of nonfacial cues, including gender (Widen & Russell, 2002), context of the situation (Carroll & Russell, 1996; Wallbott, 1988), and the expresser's nonfacial behavior. Cupchik and Poulos (1984) found that observers were 100% accurate in discerning which of four emotions was felt by expressers, when observers were given both spontaneous facial expression and the eliciting stimulus. Nonfacial information must be set aside when testing the hypothesis (found in prominent theories of emotion; Ekman, 1980; Tomkins, 1962; Izard, 1971) that the facial expression in itself conveys a distinct recognizable emotion.

Outside the laboratory, especially historically, observers are rarely limited to one view of the face. This observation suggests that our use of still photographs could have lowered the amount of agreement seen. That is, perhaps dynamic facial expressions would have yielded more consensus. Direct comparisons of dynamic versus static facial expressions have relied on a variety of methods, including computer-generated faces (Weyers, Muehlberger, Hefele, & Pauli, 2006), drawings (Wehrle, Kaiser, Schmidt, & Scherer, 2000), actors (Ambadar, Schooler, & Cohn, 2005), and hapticly presented emotional facial expressions (Lederman, Klatsky, Abromowicz, Salsman, Kitada, & Hamilton, 2007); most research examined adults, but one examined children's responses to facial expressions (Linder & Rosen, 2006). Results were mixed. In some studies, recognition was higher for dynamic than for static facial expressions, but in other studies the opposite was found (Kamachi et al., 2001). Further, when found, the advantage in recognition produced by dynamic over static faces is not large. The studies mentioned so far in this paragraph examined expressions that were posed. More relevant for present purposes are studies of dynamic spontaneous expressions. With nonselected expressions in three studies, Wagner and colleagues found that recognition from a dynamic presentation was low and similar to that obtained with still expressions (Motley & Camden, 1988; Yik et al., 1998). With stimuli selected to include only faces with visible facial expressions, Hess and Blairy found recognition between 43% and 87%. Further, the advantage of dynamic over static faces might be greater for some emotions than for others, as suggested, for example, by Keltner's (1995) description of embarrassment. Future research would thus do well to examine spontaneous and dynamic expressions.

Those who emphasize the ease and accuracy of recognition of emotion from facial expression might want to set the bar lower than consensus. They may point to our finding that Ekman's predicted labels were selected more often than random selection would predict. Indeed, Ekman (1994) pointed to chance as the bar that must be exceeded. In all, Ekman predicted 24 labels for the 20 faces (1 label each for 16 faces and 2 labels each for four blends); of these 24 predictions, 11 labels were endorsed significantly more often than chance would predict; for 13 labels, this was not the case. On the other hand, if we focus on happiness, sadness, anger, and surprise, on the grounds that these emotions were the ones studied by Ekman and Friesen (1971), we find a greater degree of agreement. For these four emotions, 35.7% of observers selected the predicted label. Further, if we set aside the prediction of a blend in three other photographs (Plates 12, 9, and 19) and instead make the predicted labels disgust, fear, and anger, respectively (i.e., the labels with the highest endorsement frequency), then the average endorsement for all six emotions studied by Ekman and Friesen (1971) would be 35.1%.

Setting the bar at chance, however, raises a number of concerns. First, Wagner et al. (1986) pointed out that one must consider the base rate of use of the emotion labels. Indeed, some labels were endorsed more frequently than others. Return to the finding that the four more traditional emotions (happiness, sadness, anger, and surprise) achieved higher agreement than the others. The most frequently used term was perplexed, but happy, angry, and sad (although not the term surprised) were next in rank order of frequency of use. Happy was the modal label for four photographs, only one of which Ekman had labeled as happy. Sad was the modal label for four photographs only one of which Ekman had labeled as sad. Angry was the modal label for one photograph, which was not the one Ekman had labeled as angry. Surprised was the modal label for only one photograph, which Ekman too had labeled as surprised.

When the bar is set at chance, one must also pay special attention to any method factors that might have inflated the amount of agreement. Indeed, here are a number of reasons to suggest that our results might overestimate the amount of agreement among observers. Our observers were college students, and college students obtain a higher degree of agreement with the predicted label than do non–college-educated participants (Wolfgang & Cohen, 1988). Our facial stimuli were selected by Ekman (1980) to represent his account of faces, and we have no data on the faces he did not select or on a representative sample of faces. In studies without selection of faces, much lower levels of agreement were found (Malatesta, Fiore, & Messina, 1987; Winkelmayer, Exline, Gottheil, Paredes, 1978). Our experimental context gave observers as much time as they wanted to examine the face and make their judgment; in the nonlaboratory world, facial expressions can be brief. We limited the response choices to 12 labels, and forced choice among a set of labels typically shows higher recognition than freely chosen labels (Russell, 1993; see also Frank & Stennett, 2001). Our design was within-subject: observers thus compared the various photographs as they progressed through the task (within-subject design typically shows higher recognition than a between-subjects design; Russell, 1994). It remains an empirical question of some importance just what amount of agreement occurs when other methods are used. Russell (1994; see also Ekman, 1994; Izard, 1994; Russell, 1995) argued that each of these methodological factors, even if individually small in their effect, cumulatively served to inflate the amount of agreement in the cross-cultural studies often cited as showing universal recognition.

This cumulative effect of different methodological factors takes on special importance when the bar for recognition is set at chance. Each of these methodological factors serves to inflate recognition by a small amount. If agreement is near the high standard of 84%, then small inflationary biases, even cumulatively, are arguably less worrisome. However, when the bar is set at chance, or even slightly above, the concern is that results are more likely to exceed

the bar because of the cumulative effect of even small methodological biases all favoring recognition.

Finally, there is another concern with setting the bar at chance. Let us suppose that endorsement of the predicted label were significantly above chance even in the nonlaboratory world and even when all methodological biases were eliminated. Such a finding would not be terribly informative, for it would mean simply that human observers do not apply emotion labels to facial expressions randomly. Randomness was never a plausible account in the first place. Results merely exceeding chance actually tell us very little about how the observer interprets a facial expression. Extremely high agreement with the predicted label convinced many researchers that observers interpret facial expressions in terms of the predicted discrete emotions. On the other hand, less agreement, such as that seen here, even when above chance, is more difficult to interpret. Perhaps observers agree on interpreting the face in some way other than in terms of discrete emotions but in a way that leads them to be better than chance in choosing the right emotion. Russell (1997) proposed that people inevitably and automatically interpret faces in terms of broad dimensions such as pleasant–unpleasant and degree of arousal. Other proposals are that people read faces in terms of social messages (Fridlund, 1997), of action tendencies (Frijda & Tcherkassof, 1997), and of cognitive appraisal dimensions (Scherer & Grandjean, in press). Once the observer has read the face in terms of general pleasantness and arousal, or of a social message, or of an action tendency, or of a dimension of cognitive appraisal (or perhaps all four), then selecting a discrete emotion label would not be a random guess but an inference from that initial interpretation. For example, a consensual interpretation of a face as conveying nothing more than an unpleasant state, or an unfriendly social message, or a tendency to avoid, or a cognitive appraisal that a goal was not met reduces the number of applicable emotion labels to the negative ones. All such accounts are consistent with nonrandom choice of emotion labels. Lack of consensus seen here on facial expressions preselected as likely to convey a specific discrete emotion invites research on such alternative ideas—research that would do well to examine spontaneous (dynamic) expressions.

References

Ambadar, Z., Schooler, J., & Cohn, J. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science, 16,* 403–410.

Barchard, K. A., & Russell, J. A. (2006). Bias in consensus scoring, with examples from ability emotional intelligence tests. *Psicothema, 18,* 49–54.

Beaupre, M. G., & Hess, U. (2005). Cross-cultural emotion recognition among Canadian ethnic groups. *Journal of Cross-Cultural Psychology, 36,* 355–370.

Buck, R. W., Savin, V. J., Miller, R. E., & Caul, W. F. (1972). Communication of affect through facial expression in humans. *Journal of Personality and Social Psychology, 23,* 362–371.

Camras, L. A., Meng Z., Ujiie T., Dharamsi S., Miyake K., Oster H., et al. (2002). Observing emotion in infants: Facial expression, body behavior, and rater judgments of responses to an expectancy-violating event. *Emotion, 2,* 179–193.

Carroll, J., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70,* 205–218.

Carroll, J., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 71,* 164–176.

Cupchik, G. C., & Poulos, C. X. (1984). Judgments of emotional intensity in self and others: The effects of stimulus context, sex, and expressivity. *Journal of Personality and Social Psychology, 46,* 431–439.

Ekman, P. (1980). *The Face of Man.* New York: Garland Publishing, Inc.

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

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115,* 268–287.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129.

Ekman, P., & Friesen, W. (1978). *The Facial Action Coding System.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion, 10,* 159–168.

Ekman, P., Friesen, W., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W., Pitcairn, T., Ricci-Bitti, P., 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.

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

Elfenbein, H. A., & Ambady, N. (2002b). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions of a different cultural group. *Current Directions in Psychological Science, 12,* 159–164.

Elfenbein, H. A., Beaupre, M., Levesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7,* 131–146.

Fernandez-Dols, J. M., Sanchez, F., Carrera, P., & Ruiz-Belda, M. A. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. *Journal of Nonverbal Behavior, 21,* 163–177.

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

Fridlund, A. (1997). The new ethology of human facial expressions. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). New York: Cambridge University Press.

Frijda, N., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 103–129). New York: Cambridge University Press.

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

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40,* 129–141.

Izard, C. (1971). *The face of emotion.* East Norwalk, CT: Appleton-Century-Crofts.

Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115,* 288–299.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception, 30,* 875–887.

Keltner, D. (1995). The signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68,* 441–454.

Keltner, D. & Bonanno, G. A. (1997). A study of laughter and dissociation:

Distinct correlates of laughter and smiling during bereavement. *Journal of Personality and Social Psychology, 73,* 687–702.

Lederman, S. J., Klatzky, R. L., Abramowicz, A., Salsman, K., Kitada, R., & Hamilton, C. (2007). Haptic recognition of static and dynamic expressions of emotion in the live face. *Psychological Science, 18,* 158–164.

Linder, J., & Rosen, L. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's Syndrome. *Journal of Autism and Developmental Disorders, 36,* 769–777.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging, 2,* 64–69.

Marsh, A., Elfenbein, H. A., & Ambady, N. (2003). Non-verbal accents: Cultural differences in facial expressions of emotion. *Psychological Science, 14,* 373–376.

Motley, M., & Camden, C. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication, 52,* 1–22.

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion, 17,* 41–51.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141.

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin, 118,* 379–391.

Russell, J. A. (1997). Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective. In J. A. Russell & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression.* pp. 295–320. New York: Cambridge University Press.

Russell, J. A., & Fernandez-Dols, J. M. (1997). *The psychology of facial expression.* New York: Cambridge University Press.

Scherer, K. R., & Ceschi, G. (2000). Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and Social Psychology Bulletin, 26,* 327–339.

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion. Affect programs or componential appraisal patterns? *Emotion, 7,* 158–171.

Scherer, K. R. & Grandjean, D. (in press). Inferences from facial expressions of emotion have many faces. Manuscript submitted for publication, University of Geneva.

Tomkins, S. (1962). *Affect, imagery, consciousness: The positive affects.* Oxford, England: Springer.

Wagner, H., MacDonald, C., & Manstead, A. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology, 50,* 737–743.

Wagner, H. L. (1990). The spontaneous facial expression of differential positive and negative emotions. *Motivation and Emotion, 14,* 27–43.

Wagner, H. L., Lewis, H., & Ramsay, S., Krediet, I. (1992). Prediction of facial displays from knowledge of norms of emotional expressiveness. *Motivation and Emotion, 16,* 347–362.

Wallbott, H. (1988). Faces in context: The relative importance of facial expression and context information in determining emotion attributions. In Scherer, K. (Ed.), *Facets of emotion* (pp. 139–160). Hillsdale, NJ: Erlbaum.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology, 78,* 105–119.

Weyers, P., Muehlberger, A., Hefele, C., & Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology, 43,* 430–453.

Widen, S., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merrill-Palmer Quarterly, 48,* 248–262.

Widen, S. & Russell, J. A. (in press). Young children's understanding of others' emotions. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed.). New York: Guilford Press.

Winkelmayer, R., Exline, R. V., Gottheil, E., & Paredes, A. (1978). The relative accuracy of US, British, and Mexican raters in judging the emotional displays of schizophrenic and normal US women. *Journal of Clinical Psychology, 34,* 600–608.

Wolfgang, A., & Cohen, M. (1988). Sensitivity of Canadians, Latin Americans, Ethiopians, and Israelis to interracial facial expressions of emotions. *International Journal of Intercultural Relations, 12,* 139–151.

Yik, M., Meng, Z., & Russell, J. A. (1998). Adults' freely produced emotion labels for babies' spontaneous facial expressions. *Cognition & Emotion, 12,* 723–730.

Received March 19, 2007
Revision received June 27, 2007
Accepted July 30, 2007 ■