COGNITION AND EMOTION, 1998, *12* (5), 723–730

Brief Report

Adults' Freely Produced Emotion Labels for Babies' Spontaneous Facial Expressions*

Michelle S.M. Yik
University of British Columbia, Canada

Zhaolan Meng
Peking University, China

James A. Russell
University of British Columbia, Canada

English-speaking Canadian, Cantonese-speaking Hong Kong Chinese, and Japanese-speaking Japanese adults were shown 13 still photographs of the facial expressions of Chinese babies subjected to various emotion-elicitation procedures. Some respondents were asked to give an emotion label of their choice for each photograph, others to judge its pleasant-unpleasant quality. Only facial expressions taken during the "happy" condition showed agreement by a majority across all three cultural samples on a specific basic emotion. Agreement on the pleasant-unpleasant quality of the baby's expressions was higher, but still varied with culture.

INTRODUCTION

"When someone feels an emotion and is not trying to disguise it, his or her face appears the same no matter who that person is or where he or she comes from" (Ekman, 1980, p. 7). If so, the face would be psychology's Rosetta stone in the study of emotion. The "established axiom" (Izard & Saxton, 1988, pp. 651–652) that certain facial expressions are easily and universally recognised biologically

*A fuller report of this study is available from the authors.

Requests for reprints should be sent to either Michelle Yik or James A. Russell, University of British Columbia, Department of Psychology, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4; e-mail: myik@unixg.ubc.ca *or* jrussell@psych.ubc.ca. Zhaolan Meng's address is Department of Psychology, Peking University, Peking 100871, China.

Preparation of this paper was facilitated by a grant from the Social Sciences and Humanities Research Council of Canada to James A. Russell. We thank Silver Chan, Carrie Hanson, Maho Harada, Rieko Ishizaki, Annie Lui, Steven So, and Paul Trapnell for their generous help in carrying out the study.

hardwired signals of discrete emotions has recently been questioned (Fridlund, 1994), even in infants (Camras, 1992; Oster, Hegley, & Nagel, 1992) for whom the "not-trying-to-disguise" qualifier can be discounted. The axiom has also been vigorously defended (Ekman, 1994), especially for infants (Izard, 1994; Izard et al., 1995).

An empirical resolution to the controversy is hindered by a lack of data. For example, there are no data concerning cross-cultural recognition of discrete emotions from *spontaneous* facial expressions at any age. Infancy is the ideal (perhaps only) age in which to search for undisguised facial signals of discrete emotions. Izard (1994) argued that as persons age, their facial expressions and their feelings become decoupled. For infants, however, feelings and faces are assumed innately and tightly coupled (Izard, 1994); indeed, production of facial emotion signals is widely assumed crucial to infant-caregiver attachment.

The present study was built on a pioneering study conducted by Meng, Yan, and Meng (1985), who created situations they thought would optimise the chances of eliciting a single specific emotion in babies 12 to 18 months of age, such as giving the baby a toy to elicit happiness or taking the toy away to elicit anger. Over 200 still photographs of babies' facial expressions were taken. Researchers trained in MAX (Izard, 1979) and FAST (Ekman, Friesen, & Tomkins, 1971) then selected those photographs that they judged most expressive of a discrete emotion. Of those so selected, 13 were available to us for the present study. All 13 showed expressive movements, but they did not necessarily pass the MAX or FAST criteria for the specific emotions predicted. For an ecologically valid study, the ideal sample of photographs would have been the original set of over 200. The 13 available were a small and nonrandom subsample, but still worthy of further study. If universally recognisable facial signals of discrete emotions occurred in the sample of over 200, the best photographs of the clearest examples should show up in this subsample of 13. On Izard's (1994) theory, a human being is always having some emotion, and therefore when a facial expression is elicited in an emotional context, especially in babies, a high proportion of those expressions can be predicted to be easily and universally recognisable as signals of discrete emotions. We examine two possibilities: that each face expressed the emotion predicted by Meng et al.'s elicitation procedure, and that it expressed another emotion.

In the present study, Canadian, Chinese, and Japanese respondents were shown Meng et al.'s (1985) 13 photographs and asked to name the emotion expressed in each. This method of free labelling is less subject to experimenter bias than the more commonly used forced-choice (Russell, 1993). Separate samples of respondents judged the pleasant-unpleasant quality of each facial expression to explore the possibility that a baby's spontaneous face communicates this quality.

METHOD

Free Labelling

Respondents. Three samples, each of 50 native speakers over the age of 15 with equal sex ratio, were recruited from English-speaking Canadians, Cantonese-speaking Hong Kong Chinese, and Japanese-speaking Japanese. Respondents in the

first two samples were approached individually in public places in their respective countries, whereas those in the Japanese sample were approached in a Canadian city where most of them were learning English at a local university; all had been living there for less than 18 months. All participants were asked to volunteer 10 minutes of their time.

Procedure. Each respondent was asked to look carefully at 9cm × 13cm black-and-white prints one after the other in a separate random order. The photographs were then collected, shuffled, and reshown, one at a time, with the question: "What emotion or mood is this baby feeling?" If the respondents gave several words, a phrase, or an event, the experimenter asked: "Can you give me the single best word to describe how this baby is feeling?" If the respondent could not respond to a particular photograph, it was reshown after the other photographs. Respondents were given as much time as they needed and all appeared to take the task seriously.

Scoring of Responses. Responses were scored by two criteria: a discrete-emotion criterion and, independently, a pleasantness-unpleasantness criterion. To score for discrete emotions, the method of Russell, Suzuki, and Ishida (1993) was used. Two native speakers (students who were not experts in emotion but blind to the facial stimuli and to the authors' hypotheses) rated every response as to whether it corresponded to one of Meng et al.'s (1985) six emotions, broadly defined. A response was scored correct if it was reasonably close to the term, but not so vague that it could apply to more than one basic emotion. Discussions ensured that the same criteria were applied in all three languages. Reliability was estimated by the percentage of agreement between the two raters in each culture: 89%, 91%, and 94% for the Canadian, Chinese, and Japanses samples, respectively. Discrepancies were resolved through discussion.

To score pleasant-unpleasant quality, one rater from each cultural group judged every free response in their own language as to whether it referred to a pleasant, to a neutral, or to an unpleasant feeling. Discussion among the three raters were used to ensure the same criteria were applied in all three languages.

Pleasantness Ratings

The method was identical to that followed in gathering the free-listing data, except that there were 24 respondents per culture and that the question posed was: "How pleasant or unpleasant is this baby feeling?" A 5-point rating scale, ranging from 1 (very unpleasant) through 3 (neutral) to 5 (very pleasant), was provided.

RESULTS

Correct According to Meng et al.'s Situations

The data were first scored on the assumption that the babies' emotion was that predicted on the basis of Meng et al.'s (1985) eliciting situation. Percentages of free labels "correct" by this discrete-emotion criterion are given in Table 1.

TABLE 1
Percentage of Agreement in Three Cultures

			Discrete Emotion Criterion				*Pleasantness Criterion*				*Direct Ratings of Pleasantness*			
Prediction	*Emotion Elicitation*	*Photo No.*	*Canadian*	*Chinese*	*Japanese*	χ^2	*Canadian*	*Chinese*	*Japanese*	χ^2	*Canadian*	*Chinese*	*Japanese*	χ^2
Happiness (positive)	A toy was shown to the baby; or a mothering person played with the baby with a toy	1	82*	88*	80*	1.25	90*	94*	88*	1.10	96*	92*	100*	2.09
		2	98*	98*	98*	0.00	98*	100*	100*	2.01	92*	100*	100*	4.11
		3	80*	84*	88*	1.19	88*	90*	96*	2.19	96*	100*	100*	2.03
		5	46*	24	22	8.34[b]	48*	34 (NEU 52)	32 (NEU 56)	3.23	67*	33 (NEU 46)	25 (NEU 42)	9.60[b]
Surprise (neutral)	A box containing a big white rat was shown to the baby	6	52*	14	52*	20.17[c]	80*	68*	72*	1.91	17 (POS 83)	58*	42 (POS 58)	8.88[b]
		7	24*	6	36	13.29[c]	72*	84*	84*	3.00	67*	92*	75*	4.50
		8	20*	2 (sad 20)	10 (sad 20)	8.54[b]	34 (NEG 44)	16 (NEG 64)	44	9.36[c]	46*	29 (NEG 63)	25 (NEG 71)	2.63
Sadness (negative)	The mothering person was playing happily with the baby, but then left and took the toy away from the baby	9	46*	30	70*	16.23[c]	100*	96*	94*	2.90	96*	92*	100*	2.09
		10	50*	22	56*	13.46[c]	96*	88*	92*	2.17	88*	100*	92*	3.01
Anger (negative)	A stranger tied the baby's hands loosely on a chair, following the departure of the mothering person; or a baby of the same age took away the toy from the baby	11	14	4 (happy 28)	10 (sad 18)	2.99	82*	36	70*	24.34[c]	96*	54*	42*	16.74[c]
		12	8	6	18	4.34	34 (NEU 58)	52*	22 (NEU 74)	9.90[c]	33 (NEU 58)	50*	54*	2.35

Disgust (negative)	The baby smelled vinegar	13	4	2 (happy 8)	4	0.41	50*	58*	46 (NEU 54)	1.49	38 (NEU 58)	50*	33 (NEU 63)	1.50
Fear (negative)	A stranger with a syringe or a big white rat appeared in front of the baby, following the departure of the mothering person	15	28*	20 (sad 20)	26	0.93	100*	94*	98*	3.60	96*	96*	96*	0.00
Mean %[a]			42.5	30.8	41.3		74.8	70.0	72.2		71.4	72.8	68.0	

Note: Number in parentheses is the percentage who agreed on an emotion other than that predicted by Meng et al. (1985). For each photograph, χ^2 was computed for agreement on the emotion predicted by Meng et al. or the pleasant-unpleasant quality predicted. The degree of freedom for all the χ^2 tests is 2. POS = positive interpretation; NEG = negative interpretation; NEU = neutral interpretation.

*Modal response was scored correct by Meng et al.'s eliciting condition. [a] Non-weighted mean % averaged across the 13 photographs. [b] $P < .05$; [c] $P < .01$.

Results of χ^2 tests showed that recognition of the correct discrete emotion varied with culture for 6 of the 13 photographs; fewer than one would be expected by chance. Results also varied between Happiness and non-Happiness conditions. For Happiness, of the 12 (4 photographs $\times$ 3 cultural groups) cases, the modal response was correct in 10; the overall mean was 74% correct. For all other emotions, of the 27 cases, modal response was correct in 9; overall mean was 23%.

Percentage correct of the pleasantness-unpleasantness responses are also given in Table 1. For the free-label data, χ^2 tests showed that recognition of the correct response varied with culture for 3 of 13 photographs; fewer than one would be expected by chance. Modal response was correct in 30 of the 39 cases. Overall average was 72% correct. For the direct-rating data, recognition varied with culture for 3 of 13 photographs; fewer than one would be expected by chance. Modal response was correct in 30 of 39 cases. Overall average was 71% correct. These two independent assessments of pleasantness largely agreed with each other. The correlations were .72, .95, and .85 for the Canadian, Chinese, and Japanese samples, respectively.

Rescoring

It is possible that an individual baby did not respond to the elicitation condition with the emotion predicted. If so, the criterion for the "correct" emotion in the previous two paragraphs might underestimate how well respondents' recognised the baby's emotion. We therefore rescored the data by allowing the modal judgement (rather than the emotion predicted by Meng et al., 1985) to determine the "correct" response. As to discrete emotion, in 33 of 39 cases, the modal judgement fell in the same emotion category predicted by Meng et al. In the six remaining cases, another emotion was modal. In these cases, percentage of agreement ranged from 8% to 28% (and are shown in parentheses in Table 1). Although these figures were (by definition) higher than those in our previous analysis, the improvement was small (and there was more opportunity to capitalise on chance). Moreover, this rescoring indicated greater cultural differences: In each case where the consensus rose by this rescoring, the modal basic emotion differed across cultures.

Pleasantness measures were similarly rescored. With the free-listing data, rescoring made no difference in 32 of 39 cases. In the remaining 7 cases, the modal response ranged from 44% to 74%. With the direct-rating data, rescoring made no difference in 30 of 39 cases. In the 9 remaining cases, the modal response ranged from 42% to 83%. As in the discrete-emotion analysis, this rescoring revealed cases where the modal response varied with culture. Nonetheless, in all cases, no pleasant interpretation was given to the photographs predicted to be unpleasant, or vice versa.

DISCUSSION: SOMETHING IS AMISS

Adults of three different cultures were shown 13 still photographs of Chinese babies' facial expressions elicited under emotional circumstances. The adults' interpretation of those expressions showed a surprising amount of disagreement

within and across cultures. When responses were scored against the criterion of a specific discrete emotion, whether as predicted or not, agreement was weak in all but the happy condition and varied as a function of culture for about half of the photographs. When free labels were scored as to pleasantness, or when pleasantness was directly rated, agreement was higher, although far from perfect.

Our data were few and preliminary, and yet they provide a sobering counterweight to predictions that derive from many theoretical writings on this topic. Taken together with the results of Camras (1992) and Oster et al. (1992), our results question the widely accepted assumption that under emotional situations, human infants routinely produce easily and universally recognisable facial signals of specific discrete emotions. Our results also reinforce the possibility that there is less agreement within and across cultures in interpreting spontaneous than posed facial expressions.

Our study, like many others, conflates the proposition that infants produce a universal facial pattern when in a specific emotional state (a test of which would require direct assessment of babies' facial actions) with the proposition that the facial behaviour they produce is easily, consensually, and universally interpreted as conveying that very emotion. This conflation precluded us from testing either proposition singly. But by assessing the amount of agreement in the interpretation of infant faces, the conjoint validity of the two propositions can be examined. This conflated method is common and fair, because "recognition" presupposes a recognisable signal. Our finding of disagreement within and across cultures indicates that something is amiss, but we cannot say whether the problem lies in what the babies produced or in how adults interpreted what the babies produced. Further, because there are no litmus tests for specific emotions, we cannot be sure of each infant's emotional state. On the assumption that emotions are frequently or even always present (Izard, 1994) and that all emotions are basic (Ekman, 1994) or patterns of basic emotions (Izard & Saxton, 1988), our scoring should have detected some recognisable signal. Again, something is amiss, but it is difficult to say what that something is.

Even our conclusion that "something is amiss" can be doubted because our sample of expressions was very small and because all expressions were seen as still photographs. Perhaps larger samples would show more recognisable facial signals, although the present results would still alert us to the ecological question of how often such signals occur. Perhaps more agreement would occur among observers actually present watching the babies in real time. Any dramatic discrepancy between our results and those produced by actual observation (or perhaps by observers shown video recordings) should surprise those researchers who routinely use still photographs, such as those provided by Ekman (1980).

Manuscript received 5 December 1996
Revised manuscript received 15 August 1997

REFERENCES

Camras, L.A. (1992). Expressive development and basic emotions. *Cognition and Emotion*, *6*, 269–283.

Ekman, P. (1980). *The face of man: Expressions of universal emotions in a New Guinea village*. New York: Garland STPM.

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

Ekman, P., Friesen, W.V., & Tomkins, S.S. (1971). Facial Affect Scoring Technique (FAST): A first validity study. *Semiotica*, *3*, 37–58.

Fridlund, A.J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.

Izard, C.E. (1979). *The maximally discriminative facial movement coding system* (MAX). Newark, DE: University of Delaware, Instructional Resources Center.

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

Izard, C.E., Fantauzzo, C.A., Castle, J.M., Haynes, O.M., Rayias, M.F., & Putnam, P.H. (1995). The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Developmental Psychology*, *31*, 997–1013.

Izard, C.E., & Saxton, P.M. (1988). Emotions. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey, & R.D. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., Vol. 1, pp. 627–676). New York: Wiley.

Meng, Z., Yan, J., & Meng, X. (1985). Queding yinger mianbu biaoqing moshi de chubu changshi [A preliminary study on facial expression patterns of infants]. *Acta Psychologica Sinica*, *1*, 55–61.

Oster, H., Hegley, D., & Nagel, L. (1992). Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas. *Developmental Psychology*, *28*, 1115–1131.

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

Russell, J.A., Suzuki, N., & Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expressions. *Motivation and Emotion*, *17*, 337–351.

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