COGNITION AND EMOTION, 1996, *10* (4), 379–407

New Findings on the Contempt Expression

Nancy Alvarado
University of California, Irvine, USA

Kimberly Jameson
University of California, San Diego, USA

Three experiments using an alternative methodology tested the meaning-relationship between facial expressions of disgust and contempt and various verbal labels. Subjects completed a language-free triad task and counterbalanced paired comparison tasks. Data were analysed using consensus analysis, multidimensional scaling, and a numerical scaling model. The two categories of facial expression were found to be distinct. Reciprocal understanding of meaning between label and expression was found for disgust but not for contempt ("annoyance" was the preferred label for the contempt expression and anger expressions were preferred as exemplars of "contempt"). These data suggest that previous results are correct for the disgust expression, but that the contempt expression has been mislabelled.

Requests for reprints should be sent to Nancy Alvarado who is now at Harvard University, Department of Psychology, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA. Requests for copies of the computer programs implementing the numerical scaling system may be sent to Kimberly Jameson at the Department of Psychology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA, 92093–0109, USA. (Please supply a PC-DOS formatted double-density computer disk.)

Preparation of this article was supported in part by National Institute of Mental Health Grant MH18931 to Paul Ekman and Robert Levenson for the NIMH Postdoctoral Training Program in Emotion Research, by National Science Foundation–North Atlantic Treaty Organisation Postdoctoral Fellowship in Science and Engineering, No. RCD-9154448 to Kimberly Jameson, and by a National Research Council Associateship awarded to Jameson. Experiment 1 was previously reported as part of a doctoral dissertation (see Alvarado, 1993).

The authors gratefully acknowledge the assistance of Louis Narens and William H. Batchelder in the completion of this work. We also thank Ed Hernandez, Chris Lombardi, Charles Lai, and the other UC Irvine Semiotics Laboratory research assistants for their help with data collection. This manuscript has benefited from comments by Paul Ekman, George Mandler, and several anonymous reviewers.

INTRODUCTION

The assumption that all subjects form the same understanding of the language-related aspects of an experiment is implicit in the use of a forced-choice response format where words are the response options. However, because a subject must always select an option, regardless of its suitability, it may be unclear how well that option corresponds to the subject's internal psychological response to the stimulus. Methods for testing the appropriateness of the response options and linking them to the stimuli are needed, especially when the paradigm is to be used in cross-cultural research. This paper addresses these difficulties with respect to the correct labelling of the so-called contempt facial expression, an issue that has attracted considerable interest in the literature due to its bearing on the universality of interpretation of facial expressions of emotion (see Ekman, 1994; Izard, 1994; Russell, 1994).

The following definitions are intended in the remainder of this article. The phrase "facial expression" refers to a specific configuration of muscle movements as portrayed in a photograph, and not to any expressive intention or to any internal emotional state. The phrase "contempt expression" does not presume its correct labelling but merely identifies the particular expression of concern using the same label used in the previous literature. The phrases "emotion term", "verbal label", and "word" are used interchangeably to refer to the verbal stimulus items presented in the tasks described later, and to the use of language by subjects to describe visual stimuli like facial expressions. Names of specific verbal items are enclosed in quotation marks.

The "Contempt" Controversy

Several researchers have asserted universality of interpretation for facial expressions of sadness, happiness, fear, disgust, anger, surprise, and more recently, contempt (Ekman, 1994). The claims for contempt are based upon cross-cultural studies using a forced-choice paradigm in which a significant majority of subjects chose the term "contempt" (or its translation) as the best label for an exemplar photo of the contempt expression (Ekman & Friesen, 1986; Ekman & Heider, 1988; Izard & Haynes, 1988; Matsumoto, 1992). As portrayed in Matsumoto and Ekman's (1988) stimuli, the muscle configuration called the "contempt expression" consists of a unilateral lip curl characterised by a single action unit (AU 14 or AU 12) using Ekman and Friesen's (1978) Facial Action Coding System (FACS).

Russell (1991a,b) published conflicting findings for the same stimuli using the same forced-choice response format, and, in contrast, used rating scale and free-listing measures. His results were offered as support for a

new theory about the influence of context and the relativity of perception of emotion in the face (see also Russell, 1994; Russell & Bullock, 1986; Russell & Fehr, 1987). Russell's divergent findings occurred when the contempt expression was presented alone or accompanied by a sad expression. When the contempt expression was presented alone, subjects selected "disgust" more frequently, and rated it as displaying more disgust than other emotions. Subjects also responded "disgust" most frequently as the "single best word to describe" the contempt expression (Russell, 1991a, p. 284). Subjects categorised the contempt expression as sad, or as portraying more sadness than other emotions, when compared with a disgust expression (see Russell, 1991a).

Generally, in emotion research, the words "contempt" and "disgust" are assumed to be close in meaning and are often collapsed into a single category encompassing both emotions. Ekman and Friesen (1988) challenged this practice with respect to the findings of Izard and Haynes (1988). Ekman and Friesen suggested that when several emotion-term labels are combined into a broader emotion category designated by a single forced-choice option, it then becomes difficult to identify which specific emotion among several the subject attributes to the stimulus photo by selecting that response option.

Methodology

Russell found that substantive results vary with changes in methodology, raising issues about the appropriateness of the methodologies used (see Ekman, 1994; Russell, 1994). Rather than critique previous work, we present a new test of the contempt expression using arguably stronger methods. For example, our approach distributes context effects due to presentation order (cf. "observer's context of judgement", Russell, 1991a, p. 150) across the experiment as random error, minimising systematic bias, thus permitting meaningful main effects to emerge. Interpretable results found under such conditions would demonstrate that there exists a stable percept beyond Russell's (1991a) finding of isolated "relativity" for single photos. Consistent with Roberts and Wedell (1994), we expect that the meaning conveyed by facial expressions can be shown to have conceptual permanence, independent of the influence of restricted experimental conditions presented in previous paradigms. Roberts and Wedell (1994) also note a possible bias due to stimulus sampling in scaling studies, especially with respect to ambiguous items. This work used paired-comparisons, the method found by Roberts and Wedell (1994) to eliminate the effects of such bias.

We use indirect scaling (e.g. triads analysed using multidimensional scaling and paired-comparisons analysed using numerical scaling), rather

than direct scaling, to avoid problems of inter-rater comparability and the untested assumption that individual subjects' ratings are interval scaled and thus can be averaged. Further, using indirect scaling, the suitability of each response option can be quantitatively evaluated, both in relation to the stimuli and in relation to the other response options. These improvements in rigour aim to resolve controversy related to methodology so that the theoretical debate can resume on a sounder empirical basis.

Overview

Working within the context of this previous research, our goal is to determine whether certain facial movements are consensually labelled by perceivers as conveying a particular meaning, rather than to test whether emotion is actually expressed or conveyed by facial expression. First, we present tests of the subjects' understanding of the facial expressions alone, then a test of the relationship between the facial expressions and a representative set of verbal labels. Experiment 1 tests whether multiple exemplars of the contempt and disgust expressions, previously labelled by Ekman and Friesen (1986), are perceived as separate and distinct in meaning, or whether they are merged into fuzzy categories with overlapping boundaries. Experiment 2 tests the suitability of a variety of alternative verbal labels for the disgust and contempt expressions, including the synonyms proposed by Izard and Haynes (1988). Results are compared to ascertain whether the conveyed meaning is the same when subjects match terms against exemplar photos as when they match photos against emotion terms. Experiment 3 introduces additional emotion photos to the scale created in Experiment 2 and tests predictions based upon it.

EXPERIMENT 1

This experiment used a language-free, visual triad task to discover whether similarity of meaning was conveyed by multiple exemplars of similarly labelled photos from the six basic emotion categories hypothesised by Ekman and Friesen (1976), plus contempt photos and neutral photos. Although photos were selected based on previous labelling studies, no verbal labels were supplied to subjects in this experiment, and the term "emotion" was never mentioned (nor any synonym for it). Subjects were free to make their choices based on infinitely many different emotions, of any type, or on no emotions at all, without influence by the examiner. Subjects were not required to state nor even be aware of the basis for their choices, and that basis was free to change from trial to trial.

It was expected that, across subjects, response consensus would only emerge from somewhat similar judgements of meaning. The application of

multidimensional scaling to this group similarity data permits us to deduce the choice factors most relevant to subjects in performing the task from any patterns evident in the plot of the scaling solution. If the disgust and contempt expressions are generally perceived to be similar in meaning, then they should be grouped together in the meaning space derived from this task.

Method

Subjects

Subjects were 28 male and female undergraduate volunteers at the University of California, Irvine. A replication under a different instruction was conducted in which subjects were 10 male and female undergraduate volunteers at the University of California, Berkeley.

Materials

Stimuli were selected from Matsumoto and Ekman's (1988) JACFEE and JACNEUF picture sets: 21 colour photos, including three photos classified as anger, disgust, contempt, fear, and surprise, two photos of happiness and sadness, plus two neutral photos. Two photos of certain categories were used to limit the total number of stimuli to 21 (the largest number possible in a balanced incomplete block design without unduly burdening subjects). Caucasian expressors were selected, balanced by sex.

Neutral photos were included as a control. We conceptualise "neutral' not as a basic emotion category (e.g. calmness, peace), but as the absence of facial movement (measurable AUs). As such, neutral items might be considered more ambiguous than other items. We anticipated that if subjects were influenced by their viewing context to make differing judgements of the emotional content of the neutral photos, that would work against a finding of consensus about meaning in the resulting similarity data.

Procedure

Photos were presented as sets of three slides projected against a surface, to groups of up to 15 subjects per session. A balanced incomplete block design (lambda = 1) consisting of 70 sets of triads was presented (Burton & Nerlove, 1976). Subjects were read instructions by an experimenter asking them to inspect each picture closely, then to select the picture that was most different from the other two, based on "the meaning conveyed by the facial expression", and specifically disregarding details like sex, age, hairstyle, and any variations introduced by the slide projec-

tors. During the replication, subjects were instructed to select the picture that was most different, based on "the emotion expressd by the person in each picture". Inspection of the resulting multidimensional scaling plots confirmed that subjects could disregard extraneous differences in the photos, and that the two instructions produced virtually the same result, as described in the following analysis.

Results

Consensus Analysis

This was applied to the picture triad responses to decide whether homogeneity of response existed among subjects. Consensus modelling is described in greater detail by Batchelder and Romney (1988, 1989). As well as providing information about the pattern of responses within a group, consensus modelling validates the generalisability of the patterns shown in a multidimensional scaling solution based on similarity data. It provides a theoretical grounding that links the results for a particular random sample to the population from which it is drawn. Further, with a finding of consensus among individual respondents, the group MDS plot based on averaged or aggregated data can be expected to be similar to the plots based on individual data.

Consensus analysis applies a probabilistic model to estimate the likelihood that each particular subject will correctly answer the set of questions presented (see Appendix). It also gives confidence estimates for the correctness of each potential response when the actual answer key is unknown. The model assumes that if responses across subjects are correlated, it is because the responses are also correlated with latent shared knowledge accessed by subjects. Patterns of correlation can then be used to determine what subjects know, and what the correct answers are likely to be.

An individual's probability of giving a correct answer is called a competence rating. Competence ratings range from -1 to 1, and are normally distributed. An individual's negative rating shows extreme disagreement with the group across the range of questions asked. Thus, idiosyncratic patterns of response are readily apparent. Mean competence is one of several measures used to evaluate whether consensus exists within a group of subjects for a particular set of questions.[1]

According to criteria established by the model's developers (Batchelder & Romney, 1988, 1989), consensus was found in the triad responses: M =

[1] Criteria for consensus are: (1) mean competence above 0.500; (2) ratio between first and second eigenvalues greater than 3:1; (3) absence of negative competence ratings in the group.

0.542, SD = 0.103, N = 28. This result shows that subjects consistently responded based on a "shared" meaning in their interpretation of the stimuli. Consensus measures for the replication were similar: M = 0.613, SD = 0.082, N = 10. Note that consensus improves and the variance decreases in the replication in which subjects were explicitly instructed to use emotion—the choice criterion used spontaneously by subjects under the ambiguous instruction.

Multidimensional Scaling (MDS)

Figure 1 shows the MDS plot of the triad task similarity data. Similarity data was obtained as follows. When one item is selected as most different in a triad, the remaining two are considered to be similar. A 21 × 21 item-by-item matrix is constructed and the column for that pair is incremented. These frequencies are considered to represent similarities. Similarities are converted to percentages by dividing the observed frequency by the number of trials in which the items were presented together. Similarity percentages are converted to distances by subtracting from 1. The similarity judgements are represented in n-dimensional space using nonmetric multidimensional scaling applied using a version of Minissa with a city-block metric (Borgatti, 1993). Stress is high (0.257) and improves slightly when a third dimension is added (0.189), but interpretability of the plot diminishes. This is comparable to stress in other MDS studies using photographed facial expressions of emotion (e.g. Roberts & Wedell, 1994; Russell, Lewicka, & Niit, 1989).

Although most widely used for determining dimensionality, multidimensional scaling also provides a convenient method for observing clustering of like items. When two items are grouped closely in an MDS plot, they may be considered cognitively similar in meaning (Green, Carmone, & Smith, 1989; Shepard, 1974; Weller & Romney, 1988). In Fig. 1, similarly labelled items appear closer to each other than to items bearing different labels, with three exceptions: (1) one fear photo appears closer to a surprise photo than to the other fear photos; (2) one sadness photo appears closer to an anger photo than to the remaining sadness photo; (3) one contempt photo appears closer to a neutral photo than to the other contempt photos. Other than these exceptions, we see grouping of items by the labels assigned to them by Matsumoto and Ekman (1988).

Discussion

With the exceptions noted earlier, the multiple exemplars of each emotion category are grouped together, supporting Ekman and Friesen's (1976) contention that these facial expressions of emotion are perceived as

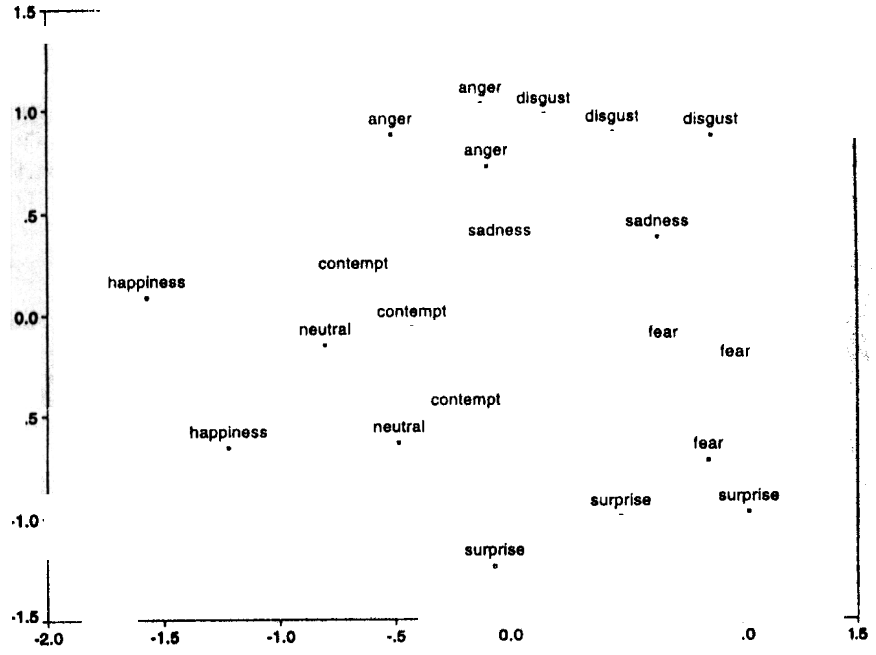

FIG. 1. MDS plot of basic emotion exemplars plus contempt. (Stress = 0.257, two dimensions.)

belonging to distinct categories of emotion. Note that verbal emotion-term labels were never supplied to subjects, and that "emotion" criteria were never mentioned during this experiment (only in the replication).

The similarity in terms of muscle movement (AUs) between the neutral and contempt expressions is reflected in their proximity in the MDS plot. From this, one might argue that this grouping occurs due to similarities in the perceived details of the faces themselves, independent of any global interpretation of emotional meaning conveyed by them. This hypothesis was tested in previous work by Alvarado (1993, submitted) with respect to similar stimuli drawn from Ekman and Friesen's Pictures of Facial Affect. That work showed that facial expressions and their corresponding verbal labels, when presented independently, are responded to in closely similar ways in a triad task, confirming that meaning itself (beyond mere perceptual similarity) is being interpreted by subjects. Also note that the groupings evident in Fig. 1 are not the result of judgements based on positive or negative evaluation, or some other simple decision rule.

Figure 1 does not resemble the traditional emotion circumplex found by Russell (1980) and Russell et al. (1989). This is because the two-dimensional solution shown in Fig. 1 represents different information than that

typically found in a plot of the first two dimensions of a multidimensional solution. Figure 1 represents *all information* in the similarity matrix fit into two dimensions, whereas a plot of the first two dimensions of a multidimensional solution disregards information related to higher dimensions.

Inspection of a three-dimensional solution for these data reveals that the relationships among items in the first two dimensions are similar to those suggested by Russell (1991a) in which contempt is close to both disgust and sadness (although not intermixed with either). However, the third dimension distinguishes contempt from both sadness and disgust. This distinctness is also evident in Fig. 1. This is consistent with the suggestion by Roberts and Wedell (1994) that the third dimension is needed to distinguish between negative emotions (i.e. anger and fear). It is also consistent with the theoretical positions taken by Osgood (1966) and Schlosberg (1954) with respect to the dimensionality of the meaning space for facial expressions of emotion. We also suggest that higher dimensionality permits relevant distinctions that are otherwise lost in two-dimensional descriptions.

Because the contempt exemplars are closest to neutral photos rather than disgust or anger photos and because sadness exemplars intervene between contempt and the anger and disgust groups, it cannot be argued that contempt is perceived as a variant of disgust or anger. Our scaling data suggest, in contrast to Russell (1991a), that the contempt photos are viewed as more similar to neutral items than to expressions of sadness. This makes intuitive sense. Although contempt and sadness are both considered negative, and both involve low arousal, these commonalities with respect to the first two dimensions are too broad to characterise the qualitative nature of the respective emotions completely. The third dimension contributes little to reduction of stress in the scaling solution, but it appears to make all the difference in distinguishing among negative low arousal emotions.

EXPERIMENT 2

Experiment 1 demonstrated that subjects consider the contempt and disgust expressions to be distinct in meaning. Experiment 2 investigates whether overlap exists in the verbal labelling applied to the two photos. Two questions are addressed: (1) which contempt or disgust photos best exemplify the emotion terms "contempt" and "disgust"; and (2) which emotion terms best describe the previously presented facial expressions of contempt and disgust? To test for reciprocity of meaning, answers to questions 1 and 2 were compared with each other. Reciprocity exists when results agree, as when, in independent tests, subjects select "contempt" as the best label for the contempt expression, and select the contempt photo as the best exemplar of the term "contempt". No a-priori linking of labels and facial

expressions is assumed by the paradigm, except the link implicit in offering any emotion term as a response option.

Method

Design

Experiment 2 consisted of two separate sets of paired comparison studies. *Study 1* consisted of a disgust condition and a contempt condition, in which different subjects were presented with paired comparisons that matched the same disgust and contempt photos pairwise against the term "disgust" (disgust condition) or the term "contempt" (contempt condition). *Study 2* consisted of eight independent conditions (one for each photo used in Study 1), in which subjects were presented with paired comparisons that matched 15 alternative labels pairwise against a single photo.

Subjects

Subjects were male and female undergraduate volunteers at the University of California, Irvine. Numbers of subjects for each condition are listed in Tables 1 and 2 (see Results). Always, the number of subjects assessed met sampling theory requirements imposed by the analysis method. For the verbal tasks in Study 2, only subjects with English as their first language were used (including subjects bilingual from early childhood). Different subjects were used in each condition and in each study.

Materials

Two types of stimuli were used: (1) facial expression photos of disgust or contempt; (2) a list of 15 emotion terms. The facial expression photos consisted of four males and four females (two in each condition), displaying either contempt or disgust (as classified by Matsumoto & Ekman, 1988), and are those used in previous contempt studies (Matsumoto, 1992; Russell, 1991a,b). The verbal items consisted of terms tested by Shaver, Schwartz, Kirson, and O'Connor (1987) for prototypicality and familiarity as descriptors of emotion (listed in Table 4), including the terms "contempt" and "disgust"; selected synonyms including "disdain" and "scorn" (as used by Izard & Haynes, 1988); also, several descriptors of anger or hostility.

The terms presented here cannot form an exhaustive set. In view of the results produced in Experiment 1, presenting terms describing a neutral

emotional state or sadness may have seemed reasonable. Proponents of alternative theoretical viewpoints (i.e. Fridlund, 1994) might similarly prefer to include terms unrelated to emotion. Such terms were not presented because they related to questions outside the scope of our inquiry. However, nothing in this methodology prevents such testing in future studies.

Procedure

The same procedure was used for both studies in Experiment 2, except as noted later. Subjects were assessed in groups of 4 to 15 using a paired comparison task presented in a counterbalanced complete pairwise design (each item was paired once with each other item). For Study 1, 8 faces were used resulting in 28 pairs. For Study 2, 15 terms were used resulting in 105 pairs. Two different random orders of pairs were used for the two conditions of Study 1 (disgust and contempt). In Study 2, different random orders were used for every four subjects.

For Study 1, a single emotion term ("disgust" or "contempt") was written in large letters on white paper and taped above a surface on which pairs of facial expression photos were projected. Subjects were read instructions that asked them to "select the photograph that shows the most contempt" (or "disgust" for that condition). For Study 2, a single photo was projected onto a surface and subjects received questionnaires listing the pairs of emotion terms. Subjects were read instructions that asked them to circle the "word" in each pair that best described the facial expression presented (a single contempt or disgust photo per session).

No definitions or other decision criteria were provided to subjects. Following the task, subjects in Study 2 were asked to review their questionnaires and mark any terms that were unfamiliar to them. Only the term "vexation" was marked by more than one subject.

Results

Consensus Analysis

This provided useful information only for Study 1, as shown in Tables 1 and 2. Conditions meeting the theory's requirements for consensus are shown by a Y (yes) in the last column. The first column of Table 2 lists photograph indices (Matsumoto & Ekman, 1988).

For Study 1, which paired eight faces against the verbal labels "contempt" and "disgust", consensus was achieved for the disgust condition but not for the contempt condition. For the contempt condition, 16 subjects

TABLE 1
Consensus Analysis Results for Experiment 2, Study 1
Single Term vs. 8 Facial Expressions

Term	*Eigenvalue Ratios*	*Mean*	*SD*		
Disgust	6.502 1.382	.676	.283	38	
Contempt					
All subjects	5.080 1.244	.042	.647	31	N
Positive subgroup	6.323 1.791	.674	95		
Negative subgroup	4.038 1.935	.573	.266	16	

* All criteria for consensus met (ratio between first and second eigenvalue 1 or more mean competence above .500); Y = yes, N = no.

TABLE 2
Consensus Analysis Results for Experiment 2, Study 2:
Single Facial Expression vs. 15 Terms

Exemplar	*Label*	*Eigenvalue Ratios*	*Mean*	*SD*	*N*	*C*[a]
ER–2C11	Contempt	3.829[b]	.498	177	5	Y
WW–1C09	Contempt	5.351 1.798	.605	146	5	Y
JH–1C10	Contempt	3.862 1.116	.438	207	16	N
KN–1C09	Contempt	2.287 1.547	.478	150	15	N
EG–1C21	Disgust	2.237 2.345	.478			
JB1–1C33	Disgust	3.740 1.147	.64	91	16	
GM–1C14	Disgust	2.407 2.525	470	31	15	N
BC–1C15	Disgust	2.948 1.632	.443	197	22	N

[a] All criteria for consensus met (see note to Table
[b] Only two factors found.

showed negative competence ratings,[2] resulting in a low mean competence for the entire group. The data from positive and negative competence subjects was partitioned, then re-analysed. As shown in Table 1, consensus was found for both subgroups. Closer analysis of the raw data revealed that the patterns of subject responses in the two subgroups were mirror images of each other; each subgroup selected the opposite response to the other subgroup for nearly all questions. No procedural or demographic explanation for this was found. The implications of this result are discussed later.

In Study 2, where 15 terms were matched against each of 8 single facial expressions (see Table 2), few of the conditions satisfied all requirements for consensus. No conclusions can be drawn from this analysis except that the domain does not meet the assumptions of the model.

Numerical Scaling Methods

The numerical scaling methods applied here (detailed in Jameson, in press), employed an algorithm that is a variant of Thurstonian scaling (Thurstone, 1927) presented by Batchelder and Bershad (1979). This model yields a continuous scale of the suitability of each term to describe the facial expression against which it was paired, or conversely, the suitability of each facial expression as an exemplar of the emotion term presented. Tables 3 and 4 present the rank orders of the rating estimates given by the numerical scaling analysis of the paired comparison data of Study 1 and Study 2, respectively. For comparison, the aggregate selection frequencies for Study 1 are shown in Table 3, expressed as a fraction of the number of pairs in which each item appeared. (Frequencies for Table 4 are available on request.) The results of the studies and the groups within studies will be discussed separately.

Scaling Photos against the "Contempt" Label

The findings for the contempt condition showed a complexity analogous to that revealed by consensus analysis. The results appear to be a compromise between the opposite judgements made by the two subgroups evident during consensus analysis by their positive and negative competence ratings. A separate scaling was carried out on these two subgroups, with several interesting results.

---

[2] According to consensus theory, negative-signed competence ratings indicate either: (1) a violation of the assumptions of the model; or (2) evidence of shared knowledge subgroups within the sample data. These options may be distinguished using the procedure applied in this study.

TABLE 3
Paired Comparison Rankings for Experiment 2, Study 1:
Single Term vs. 8 Facial Expressions

Condition	*Exemplar*	*Label*	*Ranking**	*Frequency*
Disgust		Disgust	1.128	0.782
		Disgust	1.113	0.778
		Disgust	0.842	0.711
		Disgust	0.842	0.711
		Contempt	−0.707	0.323
		Contempt	−0.977	0.256
		Contempt	−0.992	0.252
		Contempt	−1.248	0.188
Contempt				
All subjects	GM–1C14	Disgust	0.378	0.594
	KN–1C09	Contempt	0.286	0.571
	WW–1C09	Contempt	0.120	0.530
	EG–1C21	Disgust	0.028	0.507
	ER–2C11	Contempt	−0.065	0.484
	JH–1C10	Contempt	−0.157	0.461
	JB1–1C33	Disgust	−0.286	0.429
	BC–1C15	Disgust	−0.304	0.424
Positive subgroup				
	WW–1C09	Contempt	1.124	0.781
	KN–1C09	Contempt	1.086	0.771
	ER–2C11	Contempt	0.857	0.714
	JH–1C10	Contempt	0.705	0.676
	GM–1C14	Disgust	−0.324	0.419
	BC–1C15	Disgust	−1.048	0.238
	JB1–1C33	Disgust	−1.086	0.229
	EG–IC21	Disgust	−1.314	0.171
Negative subgroup				
	EG–1C21	Disgust	1.286	0.821
	GM–1C14	Disgust	1.036	0.759
	JB1–1C33	Disgust	0.464	0.616
	BC–1C15	Disgust	0.393	0.598
	KN–1C09	Contempt	−0.464	0.384
	WW–1C09	Contempt	−0.821	0.295
	ER–2C11	Contempt	−0.929	0.268
	JH–1C10	Contempt	−0.964	0.259

* All scale values shown are for iteration 2, except the negative subgroup which became stable after iteration 5.

First, the positive subgroup scaling achieved stability at the second iterative re-estimation, as did the entire group scaling, whereas the negative subgroup's scale did not converge until the fifth iteration. Secondly, although the means of the subgroup scales approximated the mean for the group, the standard deviations of the subgroup scalings were greater than

for the composite analysis (positive subgroup SD = 0.908; negative subgroup SD = 1.054; entire contempt group SD = 0.251). Thirdly, and of most interest, the rank orderings of the positive and negative subgroup scales are inverted variants of each other with the four contempt photos ranked 1 through 4 on the positive subgroup scale, and the four disgust photos ranked 1 through 4 on the negative subgroup scale. The first two differences noted previously show that the negative subgroup probably contains more noise in the choice patterns (i.e. intransitive choices) which is responsible for the delay in achieving a stable scale until iteration five.[3] The third difference suggests a strong disagreement between the two subgroups about which expression is the best exemplar of the label "contempt", and is consistent with the mirror-image consensus analysis findings described earlier.

Scaling Photos against the "Disgust" Label

The scaling of data for the disgust group is straightforward. The scale converged quickly at iteration two. Unlike the contempt group scaling, however, the four photos previously classified as disgust are the top four ranked exemplars of the label "disgust" (see Table 3, Disgust Condition).

Scaling Emotion Labels against Contempt and Disgust Photos

Scales for Experiment 2, Study 2, in which eight separate samples of subjects judged the same terms in the context of eight different photos, are shown in Table 4. All eight scales converged at the second iterative re-estimation. Thus, the emotion labels exhibited similar scaling solution characteristics no matter whether they were tested against contempt or disgust photos.

Examining the scales themselves, the label "contempt" appears to be a poor descriptor for both the contempt photos and the disgust photos. The labels "annoyance", "displeasure", and "impatience" rank in the top positions of the contempt photo context scales. On the other hand, the scales obtained for the four disgust photo contexts support the assignment of the "disgust" label to the photos. All four produced scales that identified the emotion term "disgust" as the top-ranking descriptor.

[3] Consistent with this finding, the mean competence for the positive subgroup was higher (a nonsignificant difference) than for the negative subgroup during consensus analysis. The ratio between the eigenvalues also showed that the first factor accounted for more of the variance within the positive than within the negative subgroup. This suggests greater homogeneity of response in the positive subgroup.

TABLE 4
Paired Comparison Rankings for Experiment 2, Study 2:
Single Facial Expression vs. 15 Terms

ER–2C11 *Contempt*		*WW–1C09* *Contempt*	
Annoyance	1.276	Annoyance	1.466
Impatience	1.181	Impatience	1.272
Displeasure	0.895	Displeasure	0.971
Discomfort	0.648	Dislike	0.648
Dislike	0.476	Discomfort	0.552
Resentment	0.400	Resentment	0.267
Aggravation	−0.019	Aggravation	0.191
Vexation	−0.209	Disdain	0.171
Disdain	−0.343	Exasperation	0.171
Exasperation	−0.362	Disgust	−0.152
Disgust	−0.381	Contempt	−0.971
Contempt	−0.590	Scorn	−1.048
Loathing	−0.857	Vexation	−1.067
Scorn	−0.990	Loathing	−1.219
Revulsion	−1.124	Revulsion	−1.257
JH–1C10 *Contempt*		*KN–1C09* *Contempt*	
Displeasure	1.232	Annoyance	1.467
Annoyance	0.964	Impatience	0.876
Dislike	0.750	Displeasure	0.419
Impatience	0.696	Resentment	0.381
Aggravation	0.232	Disgust	0.362
Disdain	0.143	Aggravation	0.190
Discomfort	0.125	Dislike	0.190
Exasperation	−0.054	Discomfort	0.133
Disgust	−0.089	Disdain	−0.133
Resentment	−0.250	Exasperation	−0.381
Contempt	−0.304	Revulsion	−0.419
Vexation	−0.661	Contempt	−0.571
Scorn	−0.839	Loathing	−0.629
Revulsion	−0.875	Vexation	−0.914
Loathing	−1.071	Scorn	−0.971

EG–1C21 *Disgust*		*JB–1C33* *Disgust*	
Disgust	1.467	Disgust	1.464
Revulsion	1.181	Displeasure	0.911
Loathing	0.533	Dislike	0.714
Dislike	0.324	Annoyance	0.607
Displeasure	0.267	Revulsion	0.303
Scorn	0.248	Aggravation	0.232
Aggravation	0.191	Loathing	0.071
Annoyance	0.076	Scorn	0.054
Contempt	−0.114	Disdain	−0.161
Disdain	−0.305	Discomfort	−0.179
Vexation	−0.362	Resentment	−0.518
Discomfort	−0.476	Contempt	−0.643
Resentment	−0.552	Exasperation	−0.839
Exasperation	−1.086	Impatience	−0.911
Impatience	−1.391	Vexation	−1.107
GM–1C14 *Disgust*		*BC–1C15* *Disgust*	
Disgust	1.314	Disgust	1.338
Scorn	0.819	Revulsion	0.792
Revulsion	0.800	Dislike	0.701
Loathing	0.533	Displeasure	0.662
Dislike	0.343	Loathing	0.312
Disdain	0.114	Annoyance	0.091
Aggravation	0.095	Scorn	0.013
Resentment	0.057	Disdain	−0.013
Annoyance	0.019	Discomfort	−0.026
Vexation	−0.057	Resentment	−0.208
Displeasure	−0.248	Aggravation	−0.234
Contempt	−0.400	Vexation	−0.701
Discomfort	−0.952	Exasperation	−0.753
Exasperation	−0.990	Contempt	−0.792
Impatience	−1.448	Impatience	−1.182

* All scale values shown are for iteration 2.

Results

The scaling results of Studies 1 and 2 agree in that: (1) disgust is an emotion concept that exhibits reciprocal signification (i.e. lexical descriptor *and* facial expression co-reference each other); and (2) contempt is an emotion concept that does not consistently exhibit either lexical or facial expression representation, at least for the stimuli tested here and in the previous literature.

The two paradigms (terms matched against photos vs. photos matched against terms), require two kinds of judgements that might be seen as representing two different kinds of recognition. The results of the scaling (i.e. the convergence of scales and the choice patterns in the data) suggest that subjects find it easier to judge pairs of emotion labels matched against a single photo. This might simply be attributable to differences in evaluating relatively simple word-pairs compared to evaluating the more complex photo-pairs. It may also be that subjects find it less natural to evaluate pairs of faces in combination. The "easier" judgement context is the one employed most frequently in forced-choice photo emotion identification tasks. However, each paradigm gives different information about the meaning-relation between visual and verbal items. Thus, consideration of the kind of judgement employed is important for studies that aim to construct facial expression norms.

The emotion term scaling also provides information about the relative suitability of labels proposed as synonyms for "disgust" or "contempt". Note that "disgust" was the first ranked choice for all four disgust photos. "Revulsion", "loathing", and "dislike" were highly ranked alternatives. "Contempt" is never a highly ranked descriptor for any disgust photo. Thus, "contempt" and "disgust" do not seem good synonyms in the context of a disgust expression. Synonyms for "contempt" suggested by Izard and Haynes (1988), "disdain" and "scorn", appear higher in the rankings for the disgust photos than they do for the contempt photos. This suggests that these may be better descriptors of the disgust expression, perhaps conveying a more complex, "social disgust" meaning. "Dislike" was a generic negative term appearing high in all rank order lists (hereafter referred to simply as "lists") for both types of photos.

The term "contempt" is never a highly ranked descriptor for the contempt photos either. The proposed synonyms, "disdain" and "scorn", also appear in the lower half of the lists for all four contempt photos. For three of the four contempt photos, "annoyance" is the most highly ranked descriptor, followed in most cases by "displeasure", "impatience", and "dislike", all low-intensity terms usually grouped with "anger".

About half the subjects selected disgust expressions as the best exemplars of "contempt". One explanation of this might be that disgust may

have alternative meanings: (1) physical disgust; (2) social disgust, similar to contempt. However, if half of all subjects hold a social disgust interpretation of the disgust photo when trying to match faces to the term "contempt", a similarly high proportion should consider the term "contempt" to be a suitable descriptor for the disgust photo. The emotion label numerical scales show that this is not so. Further, given the unsuitability of the term "contempt" as a descriptor of the contempt photos among nearly all subjects ("contempt" ranks 11th of 15 on two lists and 12th on two lists), a much higher proportion than half the subjects should have chosen the disgust expression as the best exemplar of "contempt" if a social disgust meaning were even remotely possible.

We believe an alternative explanation is more plausible. With respect to the contempt photos, the terms "contempt" and "disgust" are adjacent on two lists, separated by one term on a third list, and separated by two terms on the fourth list. This suggests that the inversion of scales may also result because the contempt and disgust photos are equally unsuitable as exemplars of "contempt". If the contempt photos convey neither contempt nor disgust to any great extent, as the label scales suggest, and the disgust photos are also unsuitable exemplars of "contempt" (the term "contempt" is just as low when scaled against disgust photos), then the choice among the photos may be made by chance, producing an arbitrary but consistent 50/50 split among subjects. When a choice between items is very close, so close that the selection between them may be quite difficult, factors such as random choice do enter into the decision. The consistency results because the sets of photos are clearly distinguishable. If this explanation holds, then both studies show the inadequacy of the contempt expression as an exemplar of "contempt".

EXPERIMENT 3

One advantage of the numerical scaling method presented here is the capability to incorporate efficiently new stimuli into an existing rating scale. This is an attractive feature because, in a thorough testing of labels, we might like to assess many more stimuli than are feasible to present in a single experiment. Furthermore, using an existing scale to select additional stimuli for testing is often useful, as is the ability to compare the scale values for newly incorporated items with the values of those previously scaled. Such direct comparisons have not been possible using previous methods of analysing paired comparison data. In Experiment 3 these advantages are demonstrated, as applied to questions raised in Experiments 1 and 2 earlier.

Experiment 3 is identical to the contempt condition study carried out in Experiment 2, Study 1, except for the stimuli employed. This experiment

presented four new items (hereafter called "newcomers") with four previously tested photos in a counterbalanced complete paired-comparison design. The Experiment 3 revised scale was used to predict which pictures might be preferred as exemplars of contempt if the remaining four previously tested items (not yet tested in Experiment 3) were to be subsequently matched against the newcomers. These predictions were then tested empirically in a second complete paired-comparison design and the accuracy of the predictions was evaluated.

Method

Subjects

The subjects for the first study (used to generate predictions) were 33 male and female undergraduate volunteers at the University of California, Irvine. Subjects for the second study (used to test predictions of the model) were 15 male and female undergraduate volunteers at the University of California, Berkeley.

Materials

Three types of stimulus items were used: (1) facial expression photos of disgust or contempt from Experiment 2, Study 1 above; (2) additional photos representing anger and sadness, selected from the same set (Matsumoto & Ekman, 1988); (3) a photo exemplifying a blended expression of anger and disgust created by William B. Irwin for testing in an unrelated study. Photo indices are listed in Table 5 (the blend is identified as WI-AN/DI).

Procedure

All procedures were the same as those used in Experiment 2, Study 1

Results

Consensus Analysis

This was applied to the paired-comparison data for newcomer and old items matched against the term "contempt". As in Experiment 2, consensus was not found for the entire group, so partitioned subgroups were reanalysed (subjects with positive competence ratings in one subgroup, those with negative competence ratings in another). Group consensus was found for the positive competence subgroup ($M = 0.584$), but not for the negative competence subgroup. The negative competence subgroup was partitioned

further and re-analysed, and consensus was found for second-order subgroups within the negative competence subgroup (positive $M = 0.534$ and negative $M = 0.523$).

These analyses resulted in three group of subjects for which consensus was found: (1) a 20-subject positive competence group from the first analysis; (2) a 9-subject positive competence subgroup from a second-order analysis of the negative subgroup; and (3) a 4-subject negative competence subgroup from a second-order analysis of the negative subgroup (too small for meaningful analysis and not discussed hereafter). Inspection of the estimated answer key confidence levels for the subgroups suggests that the main difference between the groups is that group (1) considered the photos portraying anger and anger/disgust to be the best exemplars of the "contempt" label, whereas group (2) considered the photos portraying contempt to be the best exemplars, followed by those portraying disgust (with anger following). This ambivalence about the correct labelling of the contempt expression is similar to that noted for Experiment 2.

Numerical Scaling Analysis

Jameson (in press) details the procedure for incorporating new stimuli into an existing scale. Table 5 presents the new scale obtained when Experiment 3 newcomer items were combined into the scale established in Experiment 2, Study 1 (contempt condition).

TABLE 5
'aired Comparison Rankings for Experiment 3, Study 1: Single Term vs. Newcomer Expressions

Condition	*Exemplar*	*Label*	*Ranking**
Contempt		Anger	
		Anger	
		Disgust	
		Anger/Disgust	
		Contempt	
		Contempt	
		Disgust	
		Contempt	
		Contempt	
		Disgust	
		Disgust	
		Sadness	

* All scale values shown are for iteration 2.

First, as in previous scalings, the scale from this combined group data converged on the second iterative re-estimation. The rank ordering of items presented in Experiment 2 was replicated in Experiment 3, and thus is unchanged from that presented in Table 3. The four newcomer items were ranked 1, 2, 4, and 12 in the revised ordering. These data show that the variants of anger photos tested (i.e. ES1-2C17, LR-1C24, WI-AN/DI) are deemed better exemplars of the "contempt" label than were several tested canonical variants of the contempt expression. The newly incorporated sadness photo (i.e. NH-1C31) is ranked at the bottom showing that it was not deemed a good exemplar of the term "contempt".

The scaling of the new stimuli into the existing scale seems consistent in that all the anger variants tested are found scaled together. As in Experiment 2, we find that typically employed contempt expressions are not the best exemplars of the "contempt" label.[4]

Predictive Capabilities of the Scaling Model

This numerical scaling method provides an ability to predict the outcomes for observations of additional pairwise choices and to predict choices for pairings of stimuli never previously paired and/or empirically assessed. Predictions of the model were made and tested by replicating Experiment 3, using the remaining four contempt and disgust photos from Experiment 2 not yet tested paired against the newcomer stimuli in Experiment 3. The testing procedures used were identical to those described for Experiment 2, Study 1 and Experiment 3, Study 1.

The method for generating predictions is described by Jameson (in press). The observed probability with which one photo was preferred to a second photo was calculated for all items empirically tested in Experiment 2, Study 1 (contempt condition) and Experiment 3. The calculations produced a 12 × 12 partial matrix of observed conditional probabilities.

[4] Note that the possibility of the convergence of the combined Experiment 2 and Experiment 3 scale at a local minimum is unlikely because an extended sequence of iterations found that the iteration 2 scale remains fundamentally unchanged after a further 1000 iterative re-estimations (i.e. varying only by a scaling constant). Further, the scale was twice derived using two different computation methods, from different starting positions, yet yields the same scaling results. That is, the scale was also produced using Experiment 3 data as the starting scale, subsequently introducing Experiment 2 items as newcomers. Both scaling approaches produced scales that are essentially identical. Both converge at the second iterative computation, both give rise to the same rank ordering of photos, and a comparison of scale values shows that the rating differences between any within-scale comparison are the same across the two scales. Finally, the two scales produced by the original and the reversed computations are perfectly correlated at $r = 1.0$ (Pearson's correlation coefficient).

The outcomes predicted by the rating system are a monotonic function of the differences in the numerical ratings of the pairs of tested stimuli. These predicted probabilities were estimated using equations given by the model.

The Goodman and Kruskal Gamma measure (see Goodman & Kruskal, 1954) was used to compare the pairwise observed choice probabilities with the model's predicted probabilities. Gamma = 0.750, and increases to Gamma = 1.00, after excluding pairs with conditional probabilities in the range 0.450 – 0.549 (4 pairs). These Gamma measures show that the rating scale and model can be employed to predict novel empirical data closely. The overall Gamma between observed and predicted pairs shows that the predictions were in accord with the empirical observations on 87.5% of the comparisons considered (percentage = 1 + Gamma/2).

GENERAL DISCUSSION

Substantive Findings for Emotion Research

Experiment 1 showed that: (1) subjects perceive previously categorised contempt and disgust photos to be dissimilar to each other in meaning, especially when compared with photos of sadness or anger; and (2) the contempt expression is not a variant of the disgust or anger expressions, but occupies its own space closer to the neutral photos (see Fig. 1). Experiment 2 showed that reciprocal signification exists for the labelling of the disgust expression, but not for the contempt expression. Among the terms provided, "disgust" is the consensual label for the disgust expression, but "annoyance" was the preferred label for three of the four contempt facial expressions. The label "contempt" was rated low on all four lists as a descriptor of the contempt expression, below "disgust" every time. However, "disgust" was also rated low on three of the four lists. Taken together, these results suggest that both "disgust" and "contempt" are poor labels for the so-called contempt expression.

More detailed analysis of Experiment 2, Study 1 showed that subjects disagree about which facial expression to match to the term "contempt", most likely because neither the contempt nor the disgust facial expression fits the term very well. A hypothesis that the disgust photos might exemplify a social disgust meaning appropriate to the term "contempt", causing some subjects to prefer disgust photos to the contempt photos, was rejected (see Discussion of Experiment 2).

Experiment 3 showed that most subjects prefer a facial expression of anger as an exemplar of the term "contempt". The division of opinion in Experiments 2 and 3 when facial expressions are matched against the term "contempt" suggests that subjects may hold two alternative meanings for the term "contempt", one related to low-level anger and consistent with

the preference for the term "annoyance" as a label, and the other more consistent with the dictionary definition of contempt. It may also be that a better exemplar of the facial expression of contempt may yield more consistent results when matched against the term "contempt". This raises the question of whether there exists an expression of annoyance distinct from contempt.

Support for Previous Findings

Our findings are consistent with Russell's (1991a,b) in that we also found the contempt expression to be problematic. Our findings do not accord with Russell's (1991a) assertions that the judgement of the contempt expression is strictly relative to viewing context (p. 161) and that there is no "direct registration of the meaning of the face" (p. 162). We found that stable and robust, consensual results were produced in a variety of changing contexts of judgement, under different presentation orders and viewing conditions (see especially Experiment 1). Therefore, we disagree with Russell's suggestion that relativity of judgement in certain limited experimental conditions means that "there is no privileged context in which the judgements to be evaluated can be made" (p. 166). We do agree that either "accurate or inaccurate judgements could be obtained . . . [by] selecting different judgement contexts" (p. 166). Russell has aptly shown how manipulating the context of judgement can produce an outcome desired by the experimenter. Therefore, we recommend methodological precautions against bias due to context, such as the use of a balanced design that presents all possible judgement contexts.

Russell's finding that the preferred label for the contempt expression is "disgust", or that "majority opinion varies with the context of judgement" (p. 166) was not supported in our studies where methodological control of context-related bias was implemented. Our finding that "annoyance" is a preferred label for the contempt expression was quite clear in four different conditions. Even the differences that were found with respect to the correct facial expression to match to the term "contempt" were consistent across contexts (compare Experiment 2, Study 1 with Experiment 3). In the viewing conditions most similar to those used by Russell (e.g. Experiment 2, Study 2), "disgust" was not the preferred label for the contempt expression (although, consistent with his findings, "disgust" was preferred to "contempt").

The forced-choice methodology does not give subjects the opportunity to select other, perhaps more preferred labels for an expression, an obvious consequence of all designs restricted to a finite number of response options. For example, subjects in all previous studies were not given the chance to select "annoyance" as a potential label for the contempt expression. Lacking

this more preferred label, subjects' responses may have been more readily influenced by the suitability of the remaining choices, as Russell suggests.

The methodological question raised by Russell (1994) with respect to choice option constraints presents a weak challenge to the universality literature unaccompanied by solid conflicting evidence. Constraint of choice options is a necessary consequence of most imaginable rigorous paradigm designs. Although this may seem to be a drawback, the advances these rigorous designs present through independent and finer-level accounting of results are also important because they permit quantitative comparisons across experiments. Thus, an ongoing cumulative assessment of tested choices becomes possible. In the long term, this investigative power allows us to work around the short-comings of constrained-choice options. Many alternatives, including direct scaling and descriptive techniques like free listing, have greater drawbacks. Use of the more extensive scaling techniques applied here, during preliminary studies, can assure that appropriate labels are presented in subsequent studies using more restricted forced-choice response formats.

Consistent with the findings of both Russell and the universality researchers, we found that when a photo of disgust is available, subjects apply the label "disgust" to it instead of the contempt photo. Anomalous findings occur when the correct labels for photos, or the correct photos for labels, are not provided. The only finding that contradicts this is the tendency of Russell's subjects to free-list the term "disgust" in response to the contempt expression. We speculate that, if the contempt expression actually conveys annoyance, then the term "disgust" may be free-listed because it is used as a synonym for "annoyance" in vernacular speech (e.g. "I am disgusted that we have to wait so long in line"). If so, subjects who have the term "annoyance" available to them should prefer it to the term "disgust", as they did in Experiment 2. This can be easily tested.

The findings of Experiment 1 are consistent with the claims of universalists for disgust. They also support the contention that contempt is perceived consistently and separately from the other "basic" emotions, and that it may therefore constitute a basic emotion itself. The findings of Experiment 1 are consistent with those claiming universality, for this culture only, using a paradigm that does not present inappropriate response options or the opportunity for context effects. However, Experiment 2 clearly shows that the preferred label for the so-called contempt expression does not appear to be "contempt".

We cannot assert that the most preferred label for the contempt expression is "annoyance", because there may exist another, more suitable but yet untested label for the expression; nor have we shown reciprocity for the term and the expression. However, we suspect that "annoyance" and translations of "annoyance" might yield higher recognition than "con-

tempt" in the kind of forced-choice paradigms used in cross-cultural studies. Obviously, we do not claim to have tested universality here.

Implications for Future Study

Elsewhere, the contempt controversy has involved seemingly incompatible findings from divergent methodologies, each with substantial flaws. Russell, in particular, has repeatedly shown that if response bias and context effects are given an opportunity to occur in a paradigm, they can produce divergent results. This cleaner methodology has reconciled previous incompatible findings by producing results that are consistent with both sides of the debate, and that suggest new directions for enquiry.

Further research is needed to learn precisely how subjects cognitively understand and define the emotion of contempt, both in relation to and independently of their definition of the emotion of disgust, and in relation to anger. The use of "disgust" as a metaphor for anger is also problematic. The nature of the basic emotion portrayed by the lip curl must be investigated. If ultimately it is an expression of annoyance, it may constitute a basic emotion in its own right. On the other hand, Ekman et al. (1991a, b) and Rosenberg and Ekman (1995) have suggested that the still photos used by previous researchers do not capture the essence of the contempt expression well enough to promote accurate labelling. Our results support this.

A clear-cut result of the present study is that the widespread practice of grouping disgust and contempt photos and/or labels into a single category appears unjustified. Verbal labels considered to be synonymous by Izard and Haynes (1988) may not be equally suitable as descriptors of these particular contempt and disgust expressions. We recommend scaling response items for similarity of meaning before they are grouped together as response options, or before they are combined as synonyms during data reduction.

What is most important, the approach demonstrated here provides a workable method for testing the meaning of labels independent of the stimuli, and the goodness of the match between them before forced-choice response options are presented to subjects. These techniques overcome difficulties cited by cultural relativists who feel that emotion cannot be studied cross-culturally because cultural assumptions are represented in the choice of labels or items.[5] When both labels and stimuli are derived

[5] Although outside the scope of this research, this methodology also presents a way of testing whether facial expressions convey emotional meaning, as opposed to other proposed content (Fridlund, 1994), because it need not suggest judgement criteria to subjects, implicitly or explicitly.

empirically and scaled in parallel but independent tasks, like those presented here, a culture-independent test of theoretical predictions can be made, even using a forced-choice response format.

Manuscript received 4 November 1994
Revised manuscript received 21 September 1995

REFERENCES

Alvarado, N. (1993). *The labeling of emotion.* Doctoral Dissertation, University of California, Irvine. University Microfilms International. Ann Arbor, MI. Microfilm no. 9323902.

Alvarado, N. (submitted). *Congruence of meaning between facial expressions of emotion and selected emotion terms.* Unpublished manuscript.

Batchelder, W.H., & Bershad, N.J. (1979). The statistical analysis of a Thurstonian model for rating chess players. *Journal of Mathematical Psychology, 19,* 39–60.

Batchelder, W.H., & Romney, A.K. (1988). Test theory without an answer key. *Psychometrika, 53,* 71–92.

Batchelder, W.H., & Romney, A.K. (1989). New results in test theory without an answer key. In E. Roskam (Ed.), *Mathematical psychology in progress* (pp. 229–248). Heidelberg: Springer.

Borgatti, S. (1993). *Reference manual: Anthropac 4.02.* Columbia, SC: Analytic Systems Inc.

Burton, M.L., & Nerlove, S.B. (1976). Balanced designs for triad tests: Two examples from English. *Social Science Research, 5,* 247–267.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 155,* 268–287.

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

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. (1988). Who knows what about contempt: A reply to Izard and Haynes. *Motivation and Emotion, 12,* 17–22.

Ekman, P., & Heider, K. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion, 12,* 303–308.

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991a). Contradictions in the study of contempt: What's it all about? Reply to Russell. *Motivation and Emotion, 15,* 293–296.

Ekman, P., O'Sullivan, M., & Matsumoto, D. (1991b). Confusions about context in the judgement of facial expression: A reply to "The contempt expression and the relativity thesis". *Motivation and Emotion, 15,* 169–176.

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

Goodman, L., & Kruskal, W. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association, 49,* 733–764.

Green, P., Carmone, Jr. F., & Smith, S. (1989). *Multidimensional scaling: Concepts and applications.* Needham Heights, MA: Allyn & Bacon.

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

Izard, C., & Haynes, O. (1988). On the form and universality of the contempt expression: A challenge to Ekman and Friesen's claim of discovery. *Motivation and Emotion, 12*, 1–16.

Jameson, K.A. (in press). Numerical scaling techniques for evaluating generative models of orthographies. *Acta Psychologica, 93*.

Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion, 16*, 363–368.

Matsumoto, D., & Ekman, P. (1988). *Japanese and Caucasian facial expressions of emotion*. Palo Alto, CA: Consulting Psychologists Press..

Osgood, C. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology, 7*, 1–30.

Roberts, J., & Wedell, D. (1994). Context effects on similarity judgments of multidimensional stimuli: Inferring the structure of the emotion space. *Journal of Experimental Social Psychology, 30*, 1–38.

Rosenberg, E., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion, 19* (2), 111–138.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Russell, J. (1991a). The contempt expression and the relativity thesis. *Motivation and Emotion, 15*, 149–168.

Russell, J. (1991b). Negative results on a reported facial expression of contempt. *Motivation and Emotion, 15*, 281–291.

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

Russell, J., & Bullock, M. (1986). Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition, 4*, 309–341.

Russell, J., & Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General, 116*, 223–237.

Russell, J., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology, 57*, 848–856.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61*, 81–88.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061–1086.

Shepard, R. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39*, 373–421.

Thurstone, L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Weller, S., & Romney, A. (1988). *Systematic data collection*. Beverly Hills, CA: Sage.

APPENDIX

The following description of the consensus model is adapted from Borgatti (1993). The model uses the following notation:

d_i	the probability that subject knows the right answer to a given question
$1-d_i$	the probability that the subject does not know the answer
L	the number of response options to a given question
1/L	the probability that the subject will guess the correct answer
$1-1/L$	the probability of guessing the incorrect answer
m_{ij}	the probability that two subjects i and j give the same answer to a given question

The parameter d_i is the subject's competence rating. It is readily calculated if the answer key is known because it is the percentage of correct questions answered minus a correction for guessing. If the answer key is not known, the parameters are estimated using the following equations:

$$m_{ij} = (d_i d_j - 1)/L \quad (1)$$

$$d_i d_j = (Lm_{ij} + 1)/L = m^*_{ij} \quad (2)$$

where m^*_{ij} is a rescaling to correct for chance guessing of the observed values m_{ij}. Equation (2) is solved for d via minimum residual factor analysis to yield a least squares estimate of the d parameter (competence rating) for each subject. Bayes theorem is then used to estimate the answer key confidence levels, given the estimated values of d.