Motivation and Emotion, Vol. 26, No. 2, June 2002 (© 2002)

Varieties of Anger: The Relation Between Emotion Terms and Components of Anger Expressions[1]

Nancy Alvarado[2,4] **and Kimberly A. Jameson**[3]

Component theory (C. Smith & H. Scott, 1997) predicts that presence of component movements (action units) alters the decoded meaning of a basic emotional expression. We tested whether the meaning of the basic expression of anger varied when different components were present in the expression. Participants were asked to label variants of anger from Ekman and Friesen's Pictures of Facial Affect *using 15 anger terms, and invariance of labeling was tested by manipulating the judgment task. Data were analyzed using consensus analysis, multidimensional scaling, and numerical scaling. Components did not result in consensus about fine distinctions in the meanings of the anger expressions. Varying the type of task strongly affected results. We believe this occurred because language elicits different categorization processes than evaluation of facial expressions nonverbally.*

KEY WORDS: facial expressions; anger; decoding; emotion; categorization; emotion naming; component theory.

Facial expressions have been widely studied because they provide a practical, if limited, test of broader theories of emotion. Studies of the decoding of facial expressions of emotion, like most studies of emotional phenomena, have depended upon the use of language to characterize emotional meaning. Recognition of this dependence on language resulted in considerable study of the emotion lexicon (Clore, Ortony, & Foss, 1987; Osgood, 1966; Shaver, Schwartz, Kirson, & O'Connor,

[1]This work was partially supported by IBM, T. J. Watson Research Center (Alvarado). The authors gratefully acknowledge the assistance of Melissa Camara, Haison Ngo, and the other UCSD undergraduates who assisted with data collection. We also thank William Irwin for his help with reliability FACS coding of facial expressions, and Louis Narens, Paul Ekman, David Matsumoto, and various anonymous reviewers for their comments on earlier drafts.
[2]Center for Brain and Cognition, University of California, San Diego, California.
[3]Department of Psychology, University of California, San Diego, California.
[4]Address all correspondence to Nancy Alvarado, Department of Psychology, University of California, San Diego, 9500 Gilman Dr., MC 0109, La Jolla, California, 92093-0109.

0146-7239/02/0600-0153/0 © 2002 Plenum Publishing Corporation

1987) and of the congruence of meaning among language, facial expression, and emotional experience (Abelson & Sermat, 1962; Alvarado, 1996; Levy, 1984; Nummenmaa, 1964; Schlosberg, 1954; Wierzbicka, 1999). Recently, the debate over universality of basic emotions, social construction, and cultural relativity of emotional experience has focused upon a series of challenges to the large body of cross-cultural studies of basic facial expressions (Ekman, 1992, 1994; Izard, 1992, 1994; Ortony & Turner, 1990; Russell, 1994; Turner & Ortony, 1992). We believe the importance of language to these challenges has been minimized or overlooked to make strong statements about the nature of emotion.

Criticisms of methodology, especially the forced-choice labeling paradigm most widely used in the cross-cultural studies, were important to these challenges to basic emotions theory (Fridlund, 1994). In several studies, simple changes in task produced different labeling results (Carroll & Russell, 1996; Russell, 1991, 1993; Russell & Bullock, 1986; Russell & Fehr, 1987). Simple manipulations also resulted in different results in studies of the lexicon alone (cf. Russell & Fehr, 1994). Nevertheless, theorists have largely attributed observed discrepancies in results across studies to the fuzziness of emotion itself (Ortony & Turner, 1990; Turner & Ortony, 1992), fuzziness in the lexicon (Russell & Fehr, 1994), or the lack of stable meaning in facial expressions (Russell & Fehr, 1987), without closely examining how language is affected by the paradigms used to test such relationships and how language in turn affects categorization and other judgments.

Much of the controversy among those studying facial expressions occurs because some researchers emphasize the subset of reliably interpreted configurations whereas others emphasize the overall unreliability of interpretation for the remaining expressions (see Ekman, 1994; Russell & Fernandez-Dols, 1997). One way of resolving this problem has been to assert a dissociation between the meaning encoded and the meaning decoded. For example, Frijda and Tcherkassof (1997) state, "that expressions manifest inner feelings probably is dogma to many researchers . . . Advance has been possible only by abandoning this dogma, recognizing that perception of emotional meaning does not necessarily imply attribution of inner feeling" (pp. 78–79). Fridlund (1997) states, "For the contemporary ethologist or behavioral ecologist, facial displays are simply messages, which influence others' behavior because vigilance for and comprehension of signals co-evolved with the signals themselves" (pp. 104).

From these statements, these theorists are proposing that facial expressions need not be expressive of an internal emotional state. We believe they assert a different source of encoding, not a dissociation between encoding and decoding. Decoding processes depend upon the information that is encoded and its utility to the observer. Unless there is reliably encoded information present in a facial expression (or emotion term), there can be no useful decoding, no matter what the proposed basis for encoding. Thus, clear thinking about the mental representation of emotional experience in language and in facial expression is needed to reason from studies of facial expression to broader emotion theory. Theories of encoding

must account for the results of decoding studies. Nor do proposed dissociations between encoding and decoding explain the existence of basic expressions, the persistent attribution of emotional states to others on the basis of facial expression, and the strong pull of attention to facial expression in both infants and adults.

The component approach has been offered as an alternative to the basic emotions approach. Component theorists assert that the basis for the encoding of facial expression is a linkage between specific facial movements and various other co-occurring phenomena of an emotional experience. Some assert that components modify the basic expressions produced by physiology. Others assert that all expressions are composed of components that occur independently and are additive in their meaning. Component theories differ primarily in their definition of what constitutes a component and what sort of information is expressed, encoded, or signaled. Smith and Scott (1997) review component approaches where encoding is based on emotional experience, including appraisals (Kaiser & Wehrle, 2001), but other theorists hypothesize encoding on the basis of social communication theories (Chovil, 1991), action tendencies, or ethology (Fridlund, 1994, 1997; Frijda & Tcherkassof, 1997).

Some theorists leave vague the link between components, language, and facial movement. Others explicitly state that fine distinctions in the meanings of facial expressions may be due to differences in the components of emotional experience, including appraisals, the environmental or social context in which the emotion occurred, and the phenomenology of the experience (Smith & Scott, 1997). Referring to these components, Carroll and Russell (1997) state, ". . . Component theory says that each AU [action unit] is due to a component" (p. 165).[5] Both Scherer (1984, 1992) and Smith (1989) hypothesize relations between specific appraisals and action units (AUs). Smith and Scott (1997) suggest that components provide ". . . a degree of fine-tuning, in which the particular components emphasized, de-emphasized, added or omitted from a given expression could provide specific information about the emotion being expressed" (p. 234).

Wierzbicka (1999) suggests that a semantics of facial expression can be developed to describe the invariant meanings of certain movements, including frowns, smiles, and eyebrow movements. She states, "The 'semantics of the human face' can rely on the same 'natural semantic metalanguage' which has been tested, over three decades, in linguistic semantics . . ." (p. 186). She supports a component approach based upon semantic meanings abstracted from both language and facial expression, and she suggests that the same semantic meanings may be conveyed by both forms of representation.

Component theory is testable only if it leads to different empirical predictions than the competing basic emotions approach. Unfortunately, encoding must be inferred from facial behavior that is subject to multiple influences. The complexity of

[5] An AU is an "action unit," as defined by Ekman and Friesen's Facial Action Coding System (Ekman & Friesen, 1978). The term refers to a momentary change in facial appearance enacted by contraction of an independent muscle or group of muscles.

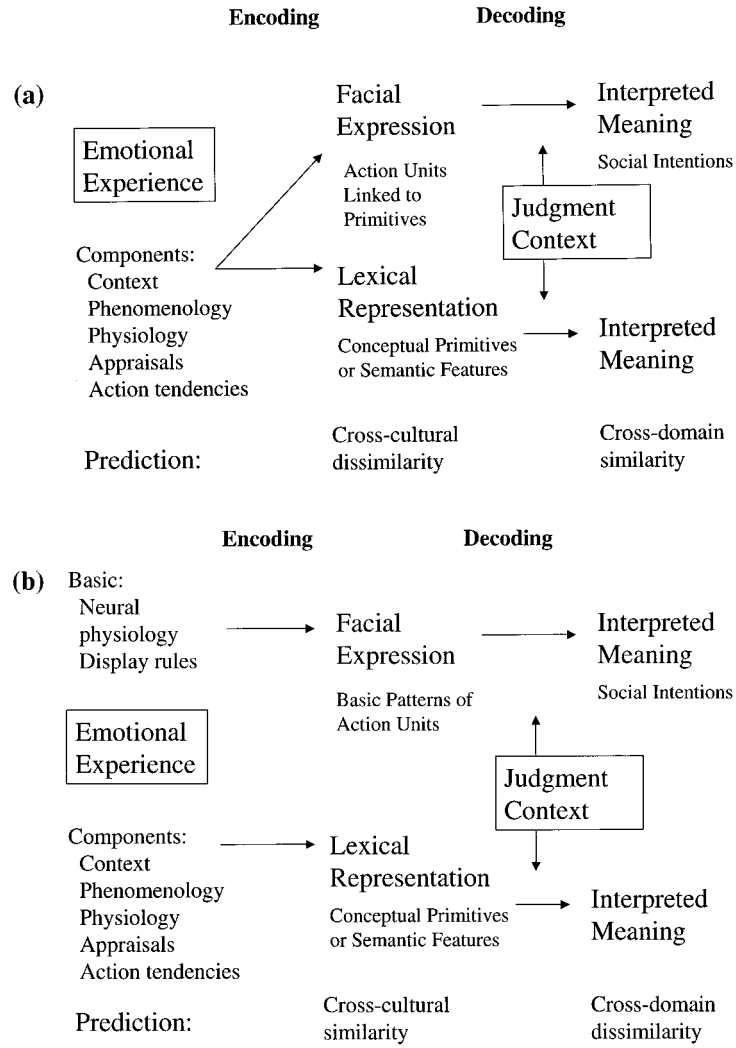

Fig. 1. Component theory predicts a similar impact on encoding facial expressions and encoding emotion terms, as shown in Fig. 1(a), whereas basic emotions theory predicts different impacts on both domains, as shown in Fig. 1(b). These differences in encoding produce different consequences for within-culture and cross-cultural agreement at decoding.

facial behavior muddies interpretation of empirical findings and makes it difficult to find clear-cut evidence in support of any theory. A result claimed as support for one side of this debate is equally likely to provide at least partial support for an alternative view. Figure 1 compares the sources of encoding proposed by component

theory (Fig. 1(a)) with those proposed by basic emotions theory (Fig. 1(b)). Note that both culture and physiology are hypothesized to play a role for both theories. Measurement of facial behavior at encoding and measurement of observer judgments at decoding are both likely to produce inconclusive results without (1) strong links back to neural physiology that also control for display rules, (2) strong manipulations of hypothesized components of emotional experience using paradigms that similarly exclude competing explanations, or (3) specific predictions about the impact of the different sources of encoding on judgments of observers. Additional complexity arises if a language-mediated task produces different results than does a nonverbal task and if the judgment context introduces variance, as Russell (1997) suggests. The research described in this paper used approach (3), which required that we additionally consider the impact of task demands on decoding judgments.

Figures 1(a) and 1(b) contrast the different sources of encoding proposed by the two theories. As shown in Fig. 1(b), the basic emotions view proposes that facial expressions are encoded as part of a coordinated pattern of physiological response that differs by type of emotion. Such expressions are modified by voluntary muscle movements that express social display rules or instrumental intentions. The universality of human physiology thus forms the basis for universally decoded meaning in such facial expressions, whereas culture influences display rules, social intentions, and the language used to describe emotion. Language encodes not simply this physiological emotional response, but also the array of components emphasized by the component theorists, resulting in a loose coupling between the decoded meanings of facial expression and description using language. As shown in Fig. 1(a), component theories propose a direct link between various components of emotional experience and specific facial movements and a direct link between such components and the language used to describe emotional experience. Because culture influences many of the initial components, it must be assumed to influence the encoding of that experience in both facial expression and language. Further, if the same components are proposed as the bases for encoding in both forms of representation, similar meanings seem more likely to emerge across these two domains of mental representation, although the encoding need not be the same.

On the basis of these differences in encoding, the two approaches make different predictions for the amount of within-culture and cross-cultural consensus about decoded meaning and the similarity of meaning between verbal labels and facial expressions. Basic emotion theory predicts cross-cultural consensus in decoding basic expressions but not other expressions, with cross-cultural variance in response to those basic expressions accounted for by display rules. Component theory predicts consensual decoding within a culture but not across cultures to the extent that the components of emotional experience vary with culture. For purposes of developing a semantics of the face, regardless of what type of component generates shared meaning (e.g., emotional experience components, social intentions, action tendencies), the existence of a shared decoded

meaning within a given culture must be demonstrable (Wierzbicka, 1999, p. 186, Item 10).

Past studies of decoding of facial expression have provided little support for a component approach because, aside from a handful of reliably interpreted combinations (closely resembling the basic expressions identified by Ekman and Friesen), little reliable decoding has been demonstrated empirically. As described by Woodworth (1938), in 1923, Boring and Tichener used Piderit's componential line drawings and found agreement for some drawings. They also found that participants were willing to attribute any emotion suggested for the other drawings, even when the parts combined were contradictory. Using the same stimuli, Buzby in 1924 and Fernberger in 1928 found that participants were willing to accept any suggested interpretation for most faces, and only objected to false labels for those expressions where an emotion was strongly indicated. Carroll and Russell (1996) produced similar results, using photos.

The finding that only certain expressions convey reliable, context invariant meaning is consistent with basic emotion theory. The finding of context dependence for the remaining expressions during decoding is problematic for Wierzbicka's assertion that the minimal meaningful units of facial behavior may be components (Wierzbicka, 1999). Component theory proposes context dependence only during encoding, not during decoding, where meaning should be stable and related to presence or absence of the specific facial movements linked to the circumstances present during encoding (Wierzbicka, 1999). Without demonstrating reliable decoding, theorists cannot assert that attending to facial expression has any social or survival-related value, much less one that depends upon componential encoding.

Smith and Scott (1997) offer empirical support for a componential approach by examining correlations between encoded frowns or smiles and affective state manipulated using directed imagery. (Wierzbicka, 1999, cites the evidence provided by Smith and Scott.) Their intention is to show that a component-based explanation is plausible on the basis of evidence provided by prior studies. However, the movements they studied are also encompassed by the larger patterns of facial expression hypothesized by basic emotion theorists. Without a direct comparison between patterns and components, this kind of evidence is inherently ambiguous.

Culture consists of shared meaning (Romney, Weller, & Batchelder, 1986). Without consensual decoding, meaning is essentially personal and idiosyncratic, and observer choices must be attributed to a combination of individual difference, judgment context specificity, and error. Further, neither social construction nor cultural relativity can be asserted for emotional experience, facial expressions, or verbal meanings without demonstrable culturally shared interpretation of meaning. Because language is the conveyor of shared meaning within a culture, it is reasonable to expect that both shared meaning and cultural differences with respect to emotional experience will be reflected in language use, both within and across cultures.

METHODOLOGICAL CONSIDERATIONS

As Wierzbicka (1999) notes, a semantic analysis of language or facial expression must focus on context-independent meaning, not the contextual interpretations that will arise during use in specific settings. Russell (1997) raises important issues about two forms of context: (1) the expresser's context—the naturally occurring circumstances in which an expression occurs and (2) the observer's context—the presentation order and judgment task in which an expression is presented for interpretation. He also asserts that the forced choice response format had a strong influence on the judgments made (Russell, 1997), especially when the only labels presented were derived from the theory being tested.

We felt it was important to avoid these methodological criticisms. Our research applied two widely used psychophysical scaling paradigms (described in Fig. 2) and associated models to inductively identify the features of our stimuli

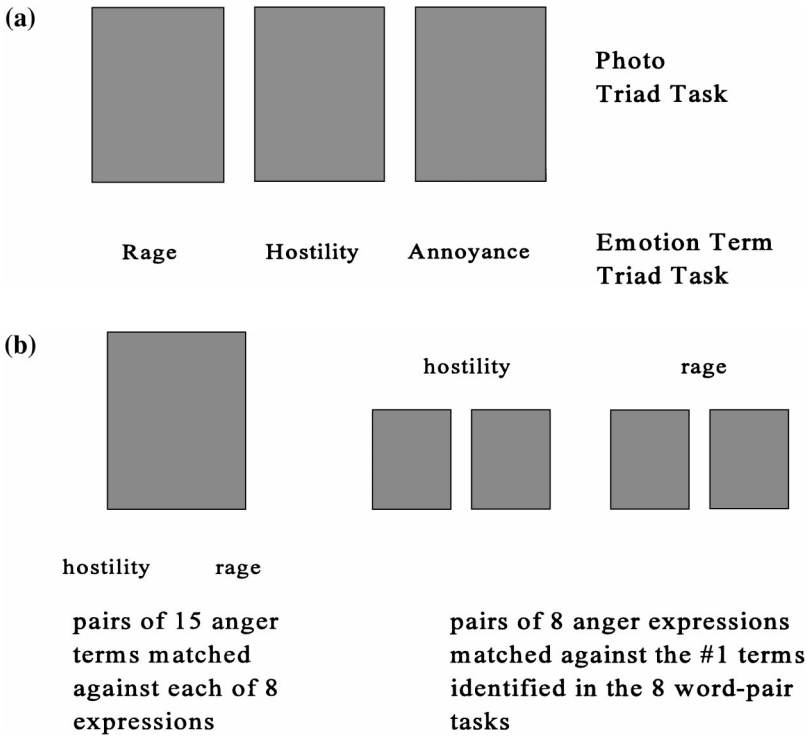

Fig. 2. The triad tasks shown in Fig. 2(a) asked participants to select the term or expression most different among three items, verbal or visual. The paired comparison tasks shown in Fig. 2(b) asked participants to select the best term to describe an exemplar facial expression or the best expression to exemplify a given emotion term (eight conditions were presented to different groups of participants).

(e.g., the AUs) most relevant to decoded meaning (Melara, 1992). To control for expresser's context, we used stimuli that provided no environmental or contextual cues beyond the expression itself (e.g., no background music, plot, scenario, dialog, tone of voice). We controlled the effects of the observer's context, using paradigms developed by Alvarado (1996) and Alvarado and Jameson (1996). As in previous studies, we systematically varied the observer's presentation order to prevent context effects and used indirect scaling to avoid rating biases. When emotion terms were offered as labels, we presented many more terms, sampled from previous studies of the lexicon (Shaver et al., 1987).

The category of "anger" was selected because it provided the greatest variability among items in both domains. Greater variability among stimuli was expected to better permit any relationship between distinct AUs and lexical meaning to emerge and thus provide a fairer test of theory. Anger expressions were taken from the *Pictures of Facial Affect* (Ekman & Friesen, 1978) because these items have been widely used in the literature, were taken under the desired controlled conditions, exclude the expresser's context, and were all identified with high reliability as anger in previous norming studies (Ekman & Friesen, 1978) despite containing very different component movements. In Shaver et al.'s study of the emotion lexicon, the category of anger contained the largest number of terms (Shaver et al., 1987). Previous studies of the structure of the anger category (Russell & Fehr, 1994; Shaver et al., 1987) provided a basis for both stimulus selection and interpretation of results.

OVERVIEW

The goals of this research were to test predictions arising from differences in theory about the encoding of emotional experience. To accomplish this, we (1) measured the similarity of decoded meaning across two semantic domains (facial expressions and emotion terms), (2) measured the extent of agreement among participants within each domain, and (3) examined the influence of simple changes in task on judgments of stimuli. We then analyzed these findings in the context of the ongoing debate between basic emotions theory and component theory.

Component theory (Smith & Scott, 1997) predicts that presence of component movements alters the decoded meaning of a basic emotional expression. Thus, expressions containing different AUs should be decoded differently, resulting in consensual selection of a different emotion term to describe them. We expected that if the meaning of a facial expression or emotion term was invariant and inherent to the stimulus, the same mapping between terms and expressions would emerge in both judgment contexts. If the component approach is correct, then consensus should emerge from all tasks, different words should be selected to describe expressions containing different components, and the same linkage between

specific emotion terms and facial expressions should be observed even when the task demands change.

STUDY 1—SIMILARITY SCALING OF STIMULI

Without conducting a separate investigation of how meaning is construed within each domain, it is difficult to interpret the results of subsequent tasks linking items across domains. Thus, two separate similarity scaling tasks were conducted, one scaling a set of 21 emotion terms, and the other scaling a set of 21 facial expressions of anger (see Fig. 2(a)). The purpose of these similarity scaling tasks was to empirically verify that participants were able to perceive the component movements present in the facial expressions used in this research, to verify that participants shared an understanding of the meanings of the anger emotion terms presented, to determine which terms or photos were considered most similar to each other in meaning (e.g., to identify synonyms), and to determine the basis for judged similarity in each domain. A triad task was used to explore similarity relationships among items.

Method

Participants

Separate groups of participants were used for the emotion term scaling task ($n = 40$) and the facial expression scaling task ($n = 33$). All participants were male or female undergraduate volunteers at the University of California, Irvine. Only native speakers of English were included.

Materials

For the emotion term scaling task, items consisted of terms tested by Shaver et al. (1987). Because there is a practical limit to the number of trials that can be presented in a single triad task without unduly burdening participants, 21 terms were selected. Terms were aggravation, anger, annoyance, bitterness, contempt, disgust, dislike, envy, exasperation, frustration, fury, grumpiness, hate, hostility, irritation, jealousy, outrage, rage, resentment, spite, and wrath. We believe the subset of anger terms presented here constitutes a reasonable and empirically justified sampling of the range of anger terms in the English lexicon. Terms found in previous research (Alvarado & Jameson, 1996) to be synonyms for disgust were omitted (e.g., revulsion, loathing), as were "agitation," "torment," "grouchiness," and "vengefulness." Agitation produced category boundary violations in Russell and Fehr's exploration

of anger terms (Russell and Fehr, 1994). Grouchiness was considered a close synonym to "grumpiness" (Alvarado, 1998). It might be argued that a term such as anger is at a different level of contrast than are more specific terms, such as "resentment" or "frustration." However, Russell and Fehr failed to find a hierarchical structure to the category of emotion terms, thus it is unclear whether any "basic" level emotion terms exist.

For the facial expression scaling task, stimulus photos were selected from Ekman and Friesen's *Pictures of Facial Affect*. Stimuli were 21 achromatic photos, consisting of all photos classified in their norming studies as anger, and four photos intended to be neutral but receiving percentages of anger responses ranging from 21 to 30%. Photos depict both male and female individuals of a variety of ages, all Caucasians. These existing photos were used in preference to creating new stimuli because extensive cross-cultural research exists demonstrating that observers perceive them as exemplifying anger, yet they do not all contain the same AUs. As discussed in the Results section, analysis showed that the photos contained sufficient variance in component AUs to adequately test predictions of component theory.

Procedure

The triad task for verbal and nonverbal items is depicted in Fig. 2(a). A balanced incomplete block design ($\lambda = 1$) consisting of 70 sets of triads was presented (Burton & Nerlove, 1976). Different participants participated in each task.

For the emotion term task, participants were given written instructions that asked them to select the word that was most different among the three on the basis of "meaning." An example was provided involving names of concrete objects. Terms were presented in sets of three words per line in a questionnaire booklet. Several different presentation orders were used, and the position of terms within the triad sets was balanced. Participants were tested in groups having 4–10 participants at a time. No photos were presented during this task.

For the facial expression task, photos were presented as sets of three slides projected against a surface, to groups of up to 10 participants per session. Participants were read instructions by an experimenter asking them to inspect each picture closely, then select the picture that was most different from the other two on the basis of "the meaning conveyed by the facial expression," and specifically disregarding details such as sex, age, hairstyle, and any variations introduced by the slide projectors. Inspection of the resulting MDS plot confirmed that these extraneous differences were not a major choice factor. No emotion terms were presented during the facial expression task, and the term *emotion* (or its synonym) was never mentioned.

Results and Discussion of Study 1

Consensus Analysis

Consensus analysis was applied to evaluate the homogeneity of response among participants. Because the consensus model may be unfamiliar to some readers, a description is provided in the Appendix. According to the criteria of the consensus model, results for the emotion term triad task came close to meeting the requirements for consensus but did not meet them: $M = .446$, $SD = .158$, $N = 40$, eigenvalue ratio (first to second factor) $= 5.11$. One participant with a negative competence score was identified. This result is lower than that generally found with verbal items, especially when scaling basic emotion terms (Moore, Romney, Hsia, & Rusch, 1999). Consensus may be reduced by the complexity of the within-category anger domain which necessarily includes terms more closely similar in meaning than are basic level emotion terms, making choices among the items more difficult.

Consensus was found in the facial expression (photo) triad responses: $M = .500$, $SD = .244$, $N = 33$, eigenvalue ratio (first to second factor) $= 3.55$. Two participants with small negative competence scores were identified. Despite these, sufficient homogeneity of response (i.e., high mean competence, high eigenvalue ratio) existed to assert that participants were likely to be drawing upon similar decoding processes in their assessments of the stimuli.

Multidimensional Scaling (MDS)

MDS was applied to the group similarity data obtained in both scaling tasks. Items in close proximity in such plots are considered to be most similar in meaning, permitting us to identify synonyms and induce the basis for the judged similarity of items (Borg & Groenen, 1997). Similarity data was obtained and converted to distances as described in previous studies (Alvarado, 1996; Alvarado & Jameson, 1996). The similarity judgments were represented in n-dimensional space using nonmetric multidimensional scaling applied using a version of Minissa with a city-block metric (Borgatti, 1993). The validity of the plots based on aggregated data was confirmed by the results of consensus analysis (see Appendix).

Figure 3 shows a two-dimensional solution for the verbal items. Unlike many two-dimensional MDS plots that show relationships among items based upon only the first two dimensions of a multidimensional solution, Fig. 3 includes information from all dimensions fit into a two-dimensional space. The increased stress of such a solution (0.302) is offset by the greater interpretability of two dimensions. Six dimensions, ultimately reducing stress to 0.080, were required to fully account for the complexity of the domain.

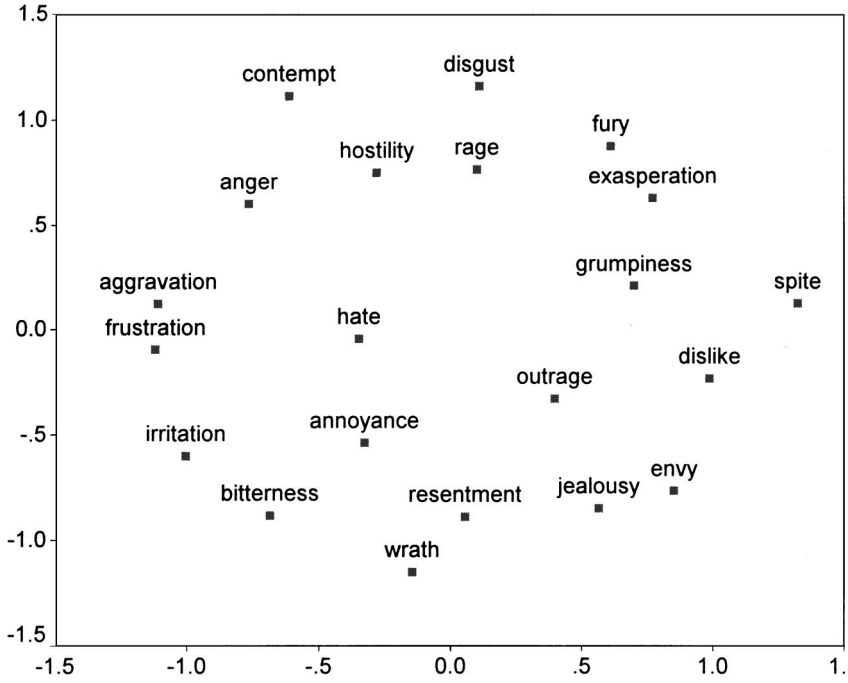

Fig. 3. Two-dimensional MDS solution for the triad similarity scaling task using 21 anger terms presented in Study 1. Note that items with similar definitions appear close to each other in this plot.

Although most widely used for determining dimensionality, multidimensional scaling also provides a convenient method for observing clustering of like items (Borg & Groenen, 1997). In Fig. 3, items that appear closer to each other in the plot may be considered rough synonyms. Intuitively, terms such as *aggravation* and *frustration* belong together and do appear together in the plot. Similarly, *jealousy* and *envy* are found together, as are *annoyance*, *resentment*, *bitterness*, and *irritation*. However, some terms such as *wrath* appear misplaced. No clear dimension of arousal, intensity, or activation emerges in the two-dimensional solution (e.g., "fury" appears near "exasperation"). Arousal or activation is important when judging across emotion categories (comparing different basic emotion terms) but appears much less important to the internal category structure of anger. Nor is anger the central term in the category; hate is most central (closest to the intersection of the *X* and *Y* axes). This suggests that anger is not a generic or basic level term despite its widespread use by theorists to name the category. An analysis of the anger lexicon is beyond the scope of this research, and so no further interpretation of the emotion term data was attempted.

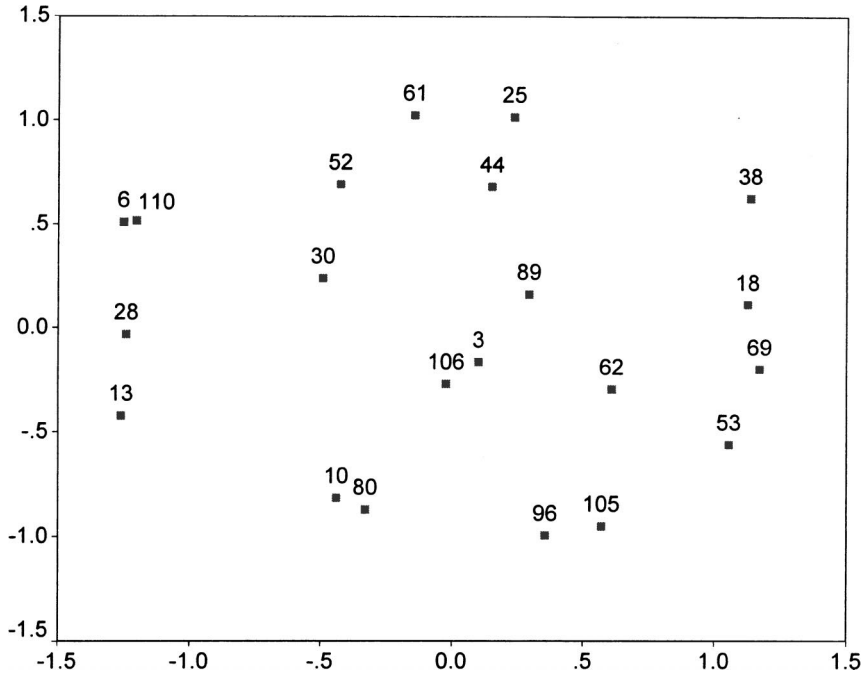

Fig. 4. Two-dimensional MDS solution for the triad similarity scaling task using 21 anger facial expressions presented in Study 1. Numbers shown are the indexes assigned to the photos by Ekman and Friesen in their *Pictures of Facial Affect*.

Figure 4 shows the MDS plot of the facial expression photo triad task similarity data. The two-dimensional MDS solution for facial expression photos shown in Fig. 4 produced lower stress than did the emotion term scaling (0.240 vs. 0.302). This is comparable to stress in other MDS studies using photographed facial expressions of emotion (e.g., Roberts & Wedell, 1994; Russell, Lewicka, & Niit, 1989).

Both domains were sampled to include maximum variability among the items within the category. Note that the domain of facial expressions appears to involve lower dimensionality, higher consensus of response, and can be depicted in a two-dimensional space with lower stress, whereas the domain of anger terms produced higher eigenvalue ratios (greater reliance on a single first factor representing shared knowledge) and less variance in competence ratings. This suggests greater constraint on meaning but more difficult choices in the verbal domain than in the visual domain. This may be the result of differences in the cognitive representations in the two domains, but it may also be the result of stimulus sampling, despite the similar approach to sampling applied to each domain.

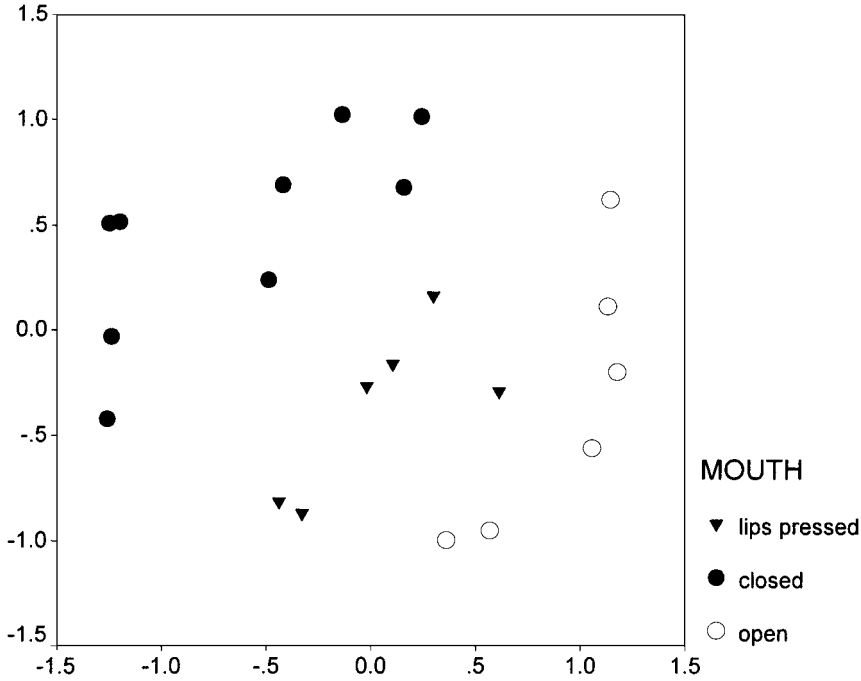

Fig. 5. Descriptions of the mouth movements present in each photo in Fig. 4 are shown in place of their item indexes. Mouths were open (AU 25 or 26), lips pressed (AU 23 or 24) or closed (no AUs). Items at the extreme left are the four "neutral" photos labeled "anger" with 20–30% frequency during norming studies. These lacked the brow lowering and eye narrowing present in the remaining anger photos.

FACS Analysis

Ekman and Friesen's Facial Action Coding System (FACS), non-baseline-corrected, was used to describe qualitative differences in the facial expression stimuli. Redundant coding of eight items by two certified coders produced an intercoder reliability above the level recommended by Ekman and Friesen. Resulting FACS AUs for each photo were superimposed onto the item positions shown in Fig. 4 to allow visual inspection of the correspondence between patterns of codes and grouping of items. This revealed that mouth AUs appeared to be most salient in distinguishing expressions, as shown in Fig. 5 (descriptive terms are used in place of AU numbers). A clear dimension ranging from closed mouth to open mouth emerges from left to right in the plot. Photos denoted as "pressed" have closed lips and exhibit AU 23 (lip tight) or AU 24 (lip press) or both. Photos in the upper center of the plot denoted as "closed" have closed lips and other anger features (such as AUs 4, 5, and 7, brow lowering, glare, and eye narrowing) but

lack AU 23 or 24. Photos at the extreme left of the plot denoted as closed have closed lips but lack other anger features. These are the four "neutral" photos that obtained 21–30% anger responses during Ekman and Friesen's norming of the picture set, most likely due to static features (such as a heavy brow line or thin lips) and a general disinclination by participants to use the "neutral" descriptor (see discussion by Rosenberg & Ekman, 1995).

It might be argued that choice factors are limited to the sources of variability existing in the photo stimulus set, or that the choice factors are dependent upon the percentage of stimuli containing such factors. Garner (1974, p. 11) suggests that sets of stimuli can be inferred by the perceiver to include items beyond those presented by an experimenter, as when a participant views a subset of letters but bases judgments upon all the letters of the alphabet. Participants bring considerable experience with facial expressions into a study, and it is possible that their judgments are based on that experience, not simply the frequencies of AUs contained in the stimuli presented. Judgments of the neutral items support this view. Despite their relative infrequency, the four items lacking basic anger AUs were clearly judged to be different from the other expressions and similar to each other, appearing together at the extreme left in Fig. 5. Other sources of variability included extent of glare (AU 5) or eye tightening (AU 7), and presence or absence of extraneous AUs such as AU 15 (lip corners depressed), AU 10 (upper lip raised), AU 17 (chin boss raised), and AU 39 (nostril compress). None of these appeared to be as strong a factor in the similarity judgments as presence/absence of AU 4+7 and mouth AUs, suggesting these other AUs had lesser salience in distinguishing the meanings of these expressions.

From the preceding triad tasks, we can conclude that participants agreed somewhat in their judgments about the meanings of emotion terms and facial expressions, that those judgments were coherent and revealed information about the choice factors used, that neutral expressions were differentiated from anger expressions (most likely by AU 4+7), and that mouth AUs appeared to be most salient in distinguishing anger expressions from each other. Thus presence or absence of various components does appear to affect the category structure for anger expressions.

STUDY 2—SCALING OF TERMS AGAINST FACIAL EXPRESSIONS

Study 2 investigated whether subtle differences in the components present in a facial expression resulted in differences in the labeling of that expression. The following questions were addressed: (1) which emotion terms are used to describe exemplars of anger with different constituent AUs and (2) do participants agree in their judgments? No a priori linking of labels and facial expressions was assumed by the paradigm.

Method

Design

Study 2 consisted of eight independent conditions (one for each of eight photos), in which eight different groups of participants were presented with paired comparisons that matched 15 alternative emotion term labels pairwise against a single photo (see Fig. 2(b)).

Participants

Participants were male and female undergraduate volunteers at the University of California, San Diego, and the University of California, Irvine. Numbers of participants for each condition are listed in Table I. Only participants with English as their first language participated (including those bilingual from early childhood).

Materials

Stimuli were a subset of the items used in Study 1 and consisted of (1) a list of 15 emotion terms and (2) eight facial expression photos of anger. The facial expression photos consisted of four males and four females selected to contain different AUs and occupying different regions of the stimulus space created in Study 1 (positions of items labeled by Ekman and Friesen's photo number are shown in Fig. 4). They included two male and two female photos with mouth open, two photos with mouth closed, two photos with lips tight or pressed. Visual stimuli are listed in Table I, also by Ekman and Friesen's index. The 15 emotion terms presented were anger, annoyance, contempt, disgust, dislike, envy, frustration, fury, grumpiness, hate, hostility, jealousy, rage, resentment, and spite.

Table I. Consensus Analysis Results for Study 2: Single Facial Expression Versus 15 Terms

PFA Index[a]	PFA Photo[a]	Term	Mean	Negatives	*N*	C[b]
MF2-7	53	Rage	.541	0	16	Yes
EM5-14	18	Frustration	.506	0	18	Yes
A1-14	3	Contempt	.479	1	15	Close
SW4-9	96	Hostility	.404	2	18	No
PE2-21	80	Anger	.276	3	15	No
WF3-1	105	Disgust	.214	9	29	No
JM5-309	44	Resentment	.168	6	16	No
GS2-8	25	Annoyance	.099	5	16	No

[a] Identifiers used in *Pictures of Facial Affect.*
[b] All criteria for consensus met (ratio between first & second eigenvalue 3:1 or more, mean competence above .500, few or no negative competence scores).

Procedure

Participants were assessed in groups of 4–15 using a paired comparison task presented in a counterbalanced complete pairwise design (each item was paired once with each other item). The 15 terms resulted in 105 pairwise combinations. Different random orders of pairs were presented for each of the conditions.

A single anger photo was projected onto a surface, and participants received questionnaires listing the pairs of emotion terms. Participants were read instructions that asked them to circle the "word" in each pair that best described the facial expression presented (a single anger photo per session). No definitions or other decision criteria were provided to participants.

Results and Discussion of Study 2

Numerical Scaling

The best descriptor for each of the eight photos was determined from the paired-comparison data in two ways: (1) Thurstonian scaling (Thurstone, 1927) to produce a rank ordering by average selection frequency and (2) a variant algorithm of Thurstonian scaling originally presented by Batchelder and Bershad (1979) and implemented by Jameson (1996) that yields an interval scale of the goodness of each term as a descriptor. Both methods produced the same ordering of items, although the latter is more informative about the difficulty of the task for participants. Complete rank orders for the 15 terms by photo are not shown because of space constraints. The single best descriptor emerging from this analysis for each photo is listed in Table I.

Consensus Analysis

Consensus analysis was applied to the paired-comparison data. Consensus varied with the constituent AUs of the photo and was present only for two open-mouthed examples of anger. Results are shown in Table I. The first column of Table I lists photograph indices (Ekman & Friesen, 1976). Conditions meeting the theory's requirements for consensus are shown by a *Yes* in the rightmost column. Four photos showed extreme disagreement among participants about the best descriptive term. Consensus analysis results (mean consensus for each photo condition) were superimposed on the facial expression MDS plot produced in Study 1 (Fig. 4). It was noted that consensus scores varied with the region of space inhabited by the photo judged. In general, photos with open mouths at the extreme right of the plot showed the highest mean competence (highest mean likelihood of producing the consensual response). Photos with lips tight or pressed in the lower center of the plot showed intermediate mean competence (but not meeting the criteria of

Table II. Correlation Between Mean Competence and Various Possible Sources of Agreement (Study 2)

	1	2	3	4	5
1. Mean competence	—				
2. Reliability[a]	.473	—			
3. Total action units	.642*	.516	—		
4. Mean intensity of action units	−.257	−.408	.242	—	
5. Percent of anger action units	.704*	.462	.640*	.322	—

[a] Based on Ekman and Friesen's reported norms for *Pictures of Facial Affect.*
* $p = .05$ (one-tailed).

the model), and photos with lips closed but not tight or pressed showed the lowest mean competence (least mean likelihood of producing the consensual response). These results confirm that component movements are important to agreement about interpreted meaning. They also suggest strongly that presence of certain AUs is related to achieving consensus about the meaning of that photo.

To explore the relationship between components and consensus, mean competence scores were compared to four other measures: (1) Ekman and Friesen's observed reliabilities for naming each photo anger in norming studies, (2) the complexity of movement in the photo as measured by the number of AUs in the face, (3) the extent of movement in the AUs present in the face as measured by the mean intensity ratings for the coded AUs, and (4) resemblance to Ekman and Friesen's hypothesized anger expression as measured by the number of AUs related to that expression present in the face. Correlations between each of these measures and mean competence are shown in Table II. Note that statistically significant correlations exist between consensus and two measures: (1) the number of AUs and (2) the resemblance to Ekman and Friesen's anger expression. The relationship between consensus and intensity of movement is nonsignificant and inverse, suggesting that extent of muscle movement is unrelated to agreement about how an expression is to be labeled. Consensus only exists for those expressions containing the largest number of anger AUs, which are generally those most closely resembling Ekman and Friesen's prototypical anger expression, and are most strongly correlated with the presence of those AUs defining the expression. As would be expected, the consensus scores track the reliabilities for the items reported by Ekman and Friesen (1978).

STUDY 3—SCALING FACIAL EXPRESSIONS AGAINST EMOTION TERMS

Study 3 investigated whether the meanings identified in Study 2 were stable with a simple change in the judgment task. Instead of presenting a photo and

asking participants to give the best name to describe it, this study presented an emotion term and asked participants to identify the best photo to exemplify it. The terms given as best descriptors in Study 2 were presented as stimuli, and different groups of participants participated. The following questions were addressed: (1) which photos are selected as the best exemplars of varieties of anger described using different emotion terms, (2) do participants agree in their judgments, and (3) is there reciprocity across the judgment contexts in Studies 2 and 3? To test for reciprocity of meaning, results from Studies 2 and 3 were compared with each other. Reciprocity existed when results agreed, as when, in independent tests, participants selected "hostility" as the best label for a particular anger expression, and selected the same anger photo as the best exemplar of the term "hostility."

Method

Design

Study 3 consisted of eight independent conditions, in which different participants were presented with paired comparisons that matched the eight anger photos used in Study 1 pairwise against a single emotion term. Each of the eight terms presented in Study 3 was identified as the best descriptor for one of the eight photos in Study 2.

Participants

Participants were male and female undergraduate volunteers at the University of California, San Diego, and the University of California, Irvine. Numbers of participants for each condition are listed in Table III. Only participants with English as their first language or bilingual from early childhood were included. Different individuals participated in each condition.

Table III. Consensus Analysis Results for Study 3: Single Emotion Term Versus 8 Facial Expressions

Presented term	Selected index[a]	PFA Photo[a]	Mean	Negatives	*N*	C[b]
Hostility	MF2-7	53	.581	1	30	Yes
Rage	MF2-7	53	.532	0	31	Yes
Disgust	MF2-7	53	.497	1	27	Close
Anger	MF2-7	53	.481	4	31	Close
Resentment	SW4-9	96	.473	2	31	Close
Frustration	MF2-7	53	.348	7	37	No
Contempt	A1-14	3	.348	5	37	No
Annoyance	MF2-7	53	.177	8	28	No

[a] Identifiers used in *Pictures of Facial Affect.*
[b] All criteria for consensus met (see Table I footnote *b*).

Table IV. Comparison of Results for Studies 2 and 3

Study 2 stimulus photo[a]	Study 2 higest ranked descriptor (Study 3 stimulus term)	Study 3 highest ranked exemplar[a]
A1-14 (3)	Contempt	A1-14 (3)
EM5-14 (18)	Frustration	MF2-7 (53)
GS2-8 (25)	Annoyance	MF2-7 (53)
JM5-03 (44)	Resentment	SW4-9 (96)
MF2-7 (53)	Rage	MF2-7 (53)
PE2-21 (80)	Anger	MF2-7 (53)
SW4-9 (96)	Hostility	MF2-7 (53)
WF3-1 (105)	Disgust	MF2-7 (53)

[a] Photos are designated by both their *Pictures of Facial Affect* participant identifier and photo number (in parentheses).

Materials

The emotion terms presented as stimuli in Study 3 are listed in Tables III and IV. They were the eight terms identified as best descriptors among the 15 anger terms presented in Study 2. The facial expression photos used were the same as those presented in Study 2 (listed in Table I).

Procedure

The procedure used in Study 2 was also used in Study 3, except as noted below. For Study 3, eight photos were used resulting in 28 pairwise combinations. Different random orders of pairs were presented for each of the conditions. A single emotion term was written in 6-in. high letters on white paper and taped above a surface on which pairs of facial expression photos were projected. Participants were read instructions that asked them to "select the photograph that is the best example of____" (whatever the term was for that condition) and "select the photograph that shows the most____."

Results and Discussion of Study 3

Consensus Analysis

Results for Study 3, which paired all eight photos against the single verbal label identified as a best descriptor for each photo in Study 3, are shown in Table III. Consensus was achieved for the conditions presenting the most extreme terms, *rage* and *hostility*, near consensus occurred for three other terms (*disgust, anger, and resentment*), and mean competence was moderate to low for the remaining terms

(contempt, frustration, and annoyance). A single photo, Item 53 (MF2-7), a photo of a woman with an open mouth and the greatest number of Ekman's hypothesized anger AUs was selected as the best exemplar for six of the eight anger terms. This declining consensus for lower activation anger terms suggests that it is easiest for participants to match a photo with many components of anger (many anger-relevant AUs) with a term signifying a great deal of anger, and much more difficult to choose the right expression to accompany a lower activation term such as annoyance. This too strongly suggests that presence or absence of movement among criterial AUs defining the basic anger expression is involved in making such judgments, rather than presence of distinctive components particular to a variant of anger. Potential candidates for such distinctive components existed in the photos, but these did not become sufficiently associated with any emotion term to evoke consensus among participants.

Numerical Scaling

In Study 3, the highest ranked anger term associated with each photo in Study 2 was presented to a group of participants as a stimulus, matched with all possible pairs of the eight anger photos used in Study 2. Numerical scaling analysis was performed as described for Study 2. The single best photo emerging from this analysis for each presented emotion term is listed in Table III. Complete rank orders for the eight photos by term are not shown because of space constraints.

Photo 53 (MF2-7) is identified as the best exemplar for six of the eight anger terms and is ranked second or third for the remaining two terms. It includes most of the AUs associated with Ekman's hypothesized anger expression and is an open-mouthed variant. It was matched with the term rage in Study 2 and produced the highest consensus of any photo. This provides additional confirmation that overall resemblance to the hypothesized basic expression, not presence or absence of individual components, mediates the linkage between anger expressions and the meanings of descriptive terms. The two exceptions are Photo 96, which emerged as the best exemplar of the term resentment, and Photo 3, which emerged as the best exemplar of contempt. Photo 96, another open-mouthed variant of anger, is highly ranked for all of the anger terms, and second choice for four terms.

Reciprocity Across Judgment Contexts

As Russell and colleagues have demonstrated, different experimental paradigms and response formats are capable of producing different results. These different results have been used to criticize the assertions of basic emotion theorists, Ekman in particular (Carroll & Russell, 1996; Fridlund, 1994, 1997; Russell & Fehr, 1987). However, such interpretations imply that our failures of reciprocity,

like the observed inconsistency across tasks of Russell and colleagues, must occur because the domain of emotion is poorly defined, fuzzy, or relative. The very assertion that results should be consistent across different types of tasks assumes that reciprocity is inherent to naming. It may not be. Similar task-dependent effects have been observed in other domains, including color naming (Jameson & Alvarado, in Press). Further, difficulties in precise definition and in application of language, independent of what is being described or named, are well known to linguists and reference philosophers (cf. Putnam, 1988).

Reciprocity assumes that when a term is selected as best describing an expression, that expression should be selected as best exemplifying that term. In other words, an experiment presenting visual stimuli and asking participants to choose or supply a name should produce the same results as an experiment presenting a single emotion term and asking participants to select the best exemplar photo from among several. This assumption that reciprocity exists across judgment contexts was explicitly tested by comparing the results of Study 2 with those of Study 3, as shown in Table IV. We found that reciprocity did not exist across rating contexts for the emotion terms and photos in this study. Only two photos and labels showed reciprocity, Photo 53 and rage, and Photo 3 and contempt. These can be loosely characterized as the best and worst anger exemplars among the photo–word pairs. Different results were found across the two contexts for all other stimuli.

Contrary to Russell's interpretation that inconsistency across judgment tasks signifies fuzziness or a context-related relativity in the use of emotion terms or perception of expressions, we believe these results suggest that participants engage in different cognitive processes in the two judgment contexts. When participants were presented with a single anger photo and 15 terms, they apparently considered the fine differences in the meanings between the terms to find the term that best matched the subtleties of the photo. When the term was presented as the stimulus, appearing alone, participants appeared to judge it merely as a generic anger term. They then selected the photo that conveyed the most anger, even when the term itself did not connote high intensities of anger (as with resentment or annoyance). The basis for evaluating how anger-like an expression is seemed to be the presence of more action units signifying anger, not the intensity or extent of movement of the action units in the photo (see Table II). Thus a different process occurred when judging terms in the context of a photo than when judging photos in the context of a term. Participants applied fine discrimination of meaning to verbal items in one context and used a more generic, basic level meaning in the other.

This use of different cognitive processes to perform different kinds of judgment tasks has also been observed with children when classifying objects. As Smith (2001) describes, children seem to categorize objects holistically when making judgments unmediated by language, but show a shape bias when asked to do a language-mediated task in which names (nouns) must be applied to categorized objects. The grammatical function of a word (e.g., whether it is a noun

or an adjective) directs a child's attention to specific stimulus features in a way that simply categorizing the objects does not. The result is different categorization judgments in the two different types of tasks, one mediated by language and one not.

Interestingly, the tendency to make fine discriminations versus more holistic basic level discriminations appears to apply to words but not to nonverbal items such as anger photos. Pairs of anger photos did not force participants to focus upon the subtleties of meaning in a single anger term. On the other hand, contrasting pairs of anger terms evoked subtleties of meaning that were then related to AUs in a single anger photo. We believe this lack of reciprocity indicates that a language-based contrast directs attention to meaningful subtleties in the word meanings whereas an equivalent nonverbal contrast does not, even when different AUs are present in the various expressions. This suggests that components in the photos do not form a semantics of emotional expression in the same way that semantic primitives or lexical features comprising the meanings of words do.

There exist observable variations in AUs in the stimulus photos to which names were systematically related in Study 2, but the judgments in that study were not being made on the basis of any shared or consensual meaning conveyed by those observable components. There was little consensus about the assignment of names for most of the anger photos, suggesting that the basis for assignment of names was idiosyncratic or personal. A different type of difficulty occurred in Study 3, where no fine discrimination of the meaning of the stimulus emotion term was made, and thus no selection of photos on the basis of component subtleties occurred. In essence, the photos lacking clear anger AUs were interpreted as ambiguous stimuli and were responded to in a variety of ways because response was forced and participants had to make some judgment, even when there was little basis for doing so. This may be the same process whereby individuals attribute meaning to similarly ambiguous facial expressions in naturalistic settings, with similar disagreement about the meanings of contextual cues (as evidenced by the low reliabilities reported in the literature).

As in the studies reported by Smith (2001), the task demands of Study 3 encouraged participants to place greater reliance on the category structure identified in Study 1 (photo triad similarity scaling). In the case of anger expressions, the category structure does not appear to consist of the subtleties or component AUs that differentiate expressions, but rather the pattern of movements shared by the anger photos that are considered definitional of anger. This potential explanation is important to the criticisms of the findings of Ekman and other basic emotion theorists. If fine discrimination emerges only in situations where it is demanded by the task, then care must be taken to evaluate whether a paradigm is capable of eliciting such discriminations. This alone can account for the discrepancies across research contexts in the literature and the strong finding of reliability only for those expressions whose AUs are most salient (e.g., those resembling Ekman and Friesen's basic expressions).

Previous research (Alvarado & Jameson, 1996) analyzing the frequency of judgment intransitivities in paired-comparison responses suggests that participants find it easier to judge pairs of emotion labels matched against a single photo than to judge photos matched against a term (see also Batchelder & Bershad, 1979; Jameson, 1996). It may be that participants find it less natural to evaluate pairs of faces in combination and thus resort to a more basic-level or holistic judgment. Or it may be that a holistic focus makes it harder to differentiate among faces in combination. The "easier" judgment context is the one employed most frequently in forced-choice photo emotion identification tasks, like those employed by Ekman. The "more difficult" task involving pairs of facial expressions is one used by Russell and colleagues to challenge such findings.

Study 1 showed that the emotion terms presented in Studies 2 and 3 were meaningful to participants. However, presence of synonyms must also be considered as an explanation for the lack of consensus when selecting a term to describe the closed mouth anger exemplars. If this explanation were correct, terms that were closely similar in meaning, as determined in Study 1, would be expected to be adjacent in the rating scale analysis. Inspection of those scales showed that synonyms occurred together in the ratings only for the highest consensus photos (96 and 53). The top rated descriptive terms for the remaining photos came from widely separate areas of the emotion term similarity space (shown in Fig. 3), suggesting that they were not perceived as closely similar in meaning. This supports an interpretation that lack of consensus when naming photos resulted from disagreement about which term was best, rather than difficulty selecting among closely similar items.

GENERAL DISCUSSION

This research investigated the contribution of component movements (AUs) to decoded meaning in anger expressions judged in several contexts. It found that participants were able to perceive subtleties in the facial expressions and in the meanings of the anger terms, but they did not agree about how the two were related. There was no consensual link between presence of component movements and the meanings of anger terms used to describe the facial expressions. This suggests that there is not a semantics of facial expression based upon component movements, and that if such movements are perceived, they are not related to the semantics of emotion naming in a straightforward manner. However, consensus in naming did appear to be strongly related to the presence of the pattern of AUs defined by Ekman and Friesen (1978) for the facial expression of anger. Our findings also suggest that the intensity of muscle contraction in an AU is unrelated to the strength of agreement about meaning of an anger expression. Expressions appear angrier when they contain more AUs related to anger, not when the AUs they do contain are present with greater observable intensity. Further, angrier faces evoke greater consensus than do less angry faces, regardless of the specific anger term selected to describe an expression.

In this forced-choice response format where participants were required to make a judgment, some label must emerge as "best" even if choices are entirely random. The amount of agreement among participants about the labeling is thus the crucial indicator of whether the choices of participants were meaningful or arbitrary. When participants attend to the components but disagree about their meanings, then responses may be internally consistent on a participant-by-participant basis but little consensus will emerge for the task. It might be suggested that sources of individual variation (e.g., verbal ability, exposure to faces, familiary with labeling processes) within our sample prevented the emergence of consensus. Consensus analysis assumes that if shared knowledge exists, different individuals within a culture will have differential access to it (Romney et al., 1986). Individual competence scores are a measure of that access. In our samples, participants were generally young (18–21 years), and thus might be expected to have reduced vocabularies compared to older participants, but they were also students at an elite university with very high admission standards and thus may be expected to have higher verbal competence than the general population. Because we did not measure verbal skills explicitly, we do not know the extent to which such demographic factors affected mean consensus in the tasks using verbal items in our study. There is undoubtedly some impact, but we believe that demographics alone cannot explain our findings. Failure of the goodness-of-fit measures of the model to demonstrate consensus implies more disagreement than simple variance in accessing shared meaning. Further, the existence of consensus for some conditions but not others must still be explained given that all samples were recruited from the same student population, in the same manner, and the same set of verbal stimuli was used across paired-comparison tasks.

It is also possible that components are related to the circumstances of encoding as hypothesized, but are only interpretable and meaningful for judgments about individuals whose behavior is well known through previous experience. In other words, it is possible that the meanings encoded by components are personal and idiosyncratic to the expresser rather than shared. If so, component-based decoding would be expected to be inaccessible to strangers. Or perhaps the encoding of facial expression can only be consensually decoded from dynamic stimuli because it resides in transitions between expressions or some other characteristic not present in a still photo. Such explanations would account for the strong appeal of component-based explanations, the confidence people feel in their ability to decode facial expression, and the consistent lack of success empirically demonstrating shared meaning for componential decoding. However, while intuitively appealing there is no evidence in support of such explanations because they have not yet been explored empirically.

We believe that on the basis of our findings, the challenges to Ekman and Friesen's universality studies must be reevaluated (e.g., Fridlund, 1997; Russell, 1994, 1997), as must the currently proposed component approaches. It seems most likely to us that the inconsistencies noted by Russell arise not from fuzziness in

the domain of emotion, cultural relativity, or a failure of universality, but from differences in the cognitive processes applied to the stimuli, arising from the different task demands in the various paradigms contrasted. We believe this different mapping of terms to referents results from generalized categorization and naming processes also important in domains beyond emotion research (for a color naming example, see Jameson & Alvarado, in Press). As discussed by Smith (2001), categorization may rely upon holistic judgments when language is not involved, but focus upon specific attributes of stimuli when words are used to label the stimuli being categorized. The acquisition of a vocabulary for naming thus directs attention to specific aspects of the objects to be named (see also Soja, 1994).

The lack of reciprocity in Studies 2 and 3 and the reliance upon resemblance to a prototypical anger expression (as defined by Ekman & Friesen, 1978) show that the cognitive processes applied to evaluate facial expressions are different than those used to evaluate emotion terms. Results can be expected to differ when tasks shift attention from the words to the faces, and vice versa. This difference in response to facial expressions judged as pairs, compared to emotion terms judged as pairs, strongly suggests that the representation of meaning in the two sets of stimuli is not the same. This finding is problematic for Wierzbicka's suggestion that both domains may rely upon metasemantic primitives comprising a semantics of facial expression comparable to the semantics of language.

Although the research reported here did not vary demands for speeded response, such demands may also contribute to a divergence of findings across previous studies. Speeded response forces a reliance on salient cues that impose a category structure on the domain, whereas nonspeeded response permits fine discrimination (Callaghan, 1992; Garner, 1970; Miller, 1956/1994). Thus tasks explicitly or implicitly demanding speeded response may encourage participants to recognize and respond in terms of basic expressions, not fine discriminations in facial expression stimuli. We think it likely that recognition of the patterns constituting basic expressions evolved to meet a need for quick response in naturalistic settings.

Component theories have considerable intuitive appeal, especially when researchers have historically had such difficulty obtaining reliable responses across experimental contexts. The research presented here, exploring the relationship between components in anger expressions and subtle differences in the meanings of anger terms, suggests that componential approaches will not produce any higher reliabilities today, using photos, than were obtained early last century. To be plausible, component theorists must produce empirical findings supporting their speculations. Simply on the basis of the lack of consensual meaning attributable to different components, strong versions of component theory appear incompatible with results like those presented here. In particular, we found little support for the idea that a set of phoneme-like (or visime) minimal units of meaning in the face can be identified along with a set of rules for their combination.

It might be argued that the meanings of components are linked to aspects of emotional experience, not to emotion terms, thus language-related tasks are unsuitable for measuring the meaning of components in facial expressions. However, such a claim implies that the emotion terms are not themselves descriptive of emotional experience in a systematic way (at least within a given culture). Such a claim is inconsistent with the purposes of language. We believe that those proposing strong versions of component theory must not simply criticize the universality theories. They must also account for results like those presented here, a set of studies that gave a fair test of the impact on meaning of components, and found little support for a component approach.

APPENDIX

A formal description of the consensus model has been provided by Batchelder and Romney (1988, 1989). In addition to evaluating the extent of agreement among participants, an important use of consensus modeling in this research was to validate the solutions produced by multidimensional scaling. The axioms of the formal consensus model provide a theoretical grounding that links the results for a particular random sample to the population from which it was drawn. Further, with a finding of consensus among individual respondents, the group MDS plot based upon averaged or aggregated data can be expected to be closely similar to the plots based upon individual data. This validates the generalizability of the patterns shown in a multidimensional scaling solution based upon similarity data.

Consensus analysis is a formal computational model that uses the pattern of responses within a data set to (a) predict the likelihood of correct response for each participant (called the competence rating), (b) provide an estimate of the homogeneity of response among participants (the mean competence), and (c) provide confidence estimates for the correctness of each potential response to a set of questions (the "correct" response is the consensual response). Although this model also makes certain assumptions, it incorporates goodness-of-fit measures that permit an analysis of the extent to which those assumptions have been met. In this research, consensus modeling is used as a heuristic to evaluate the amount of agreement among participants across their responses (as opposed to calculating reliability for separate items). No assumptions are being made about the potential sources of agreement.

The measures used to evaluate results are (1) individual competence scores, (2) mean competence, (3) eigenvalues produced during the minimum residual factor analysis used to estimate the solution to the model's equations, and (4) answer key confidence estimates. Competence scores range from -1.00 to 1.00 and are maximum-likelihood parameter estimates. They are best understood as estimated probabilities of producing the consensual response rather than correlation coefficients. A negative competence score indicates extreme and consistent disagreement

with the group across the entire set of responses. Answer key confidence estimates were used to evaluate the sufficiency of the sample sizes, which met model sampling requirements in all studies reported in this paper (Batchelder & Romney, 1989).

Batchelder and Romney (1988, 1989), Romney et al. (1986), and Weller and Romney (1988) suggest three criteria for judging whether consensus exists in participant responses to questions about a domain: (1) eigenvalues showing a single dominant factor (a ratio greater than 3:1 between the first and second factors), (2) a mean competence greater than 0.500, and (3) absence or near absence of negative competence scores in the group of participants. Although failure to meet these criteria does not necessarily rule out consensus, it can indicate a poor fit between the data and the model. Establishing these criteria in advance of study and in accordance with other researchers permits a more objective evaluation of whether our results reveal homogeneity of response. This addresses Russell's complaint that researchers have no way to decide how much agreement short of 100% is sufficient among participants in order to say that they agree (Russell & Fehr, 1994).

REFERENCES

Abelson, R., & Sermat, V. (1962). Multidimensional scaling of facial expressions. *Journal of Experimental Psychology, 63,* 546–554.

Alvarado, N. (1996). Congruence of meaning between facial expressions of emotion and selected emotion terms. *Motivation and Emotion, 20,* 33–61.

Alvarado, N. (1998). A reconsideration of the structure of the emotion lexicon. *Motivation and Emotion, 22,* 329–344.

Alvarado, N., & Jameson, K. (1996). New findings on the contempt expression. *Cognition and Emotion, 10,* 379–407.

Batchelder, W. H., & Bershad, N. J. (1979). The statistical analysis of a Thurstonian model for rating chess players. *Journal of Mathematical Psychology, 19*(1), 39–60.

Batchelder, W. H., & Romney, A. K. (1988). Test theory without an answer key. *Psychometrika, 53*(1), 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, Germany: Springer.

Borg, I., & Gorenen, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.

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

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

Callaghan, T. (1992). Structure in the process of seeing. In B. Burns (Ed.), *Percepts, concepts and categories: The presentation and processing of information* (pp. 149–171). Amsterdam: North-Holland.

Carroll, J., & Russell, J. (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. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology, 72,* 164–176.

Chovil, N. (1991). Social determinants of facial displays. *Journal of Nonverbal Behavior, 15,* 141–154.

Clore, G., Ortony, A., & Foss, M. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology, 53,* 751–766.
Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200.
Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115*(2), 268–287.
Ekman, P., & Friesen, W. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1,* 56–75.
Ekman, P., & Friesen, W. (1978). *Facial action coding system manual*. Palo Alto, CA: Consulting Psychologists Press.
Fridlund, A. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.
Fridlund, A. (1997). The new ethology of human facial expressions. In J. Russell & J. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 103–129). Paris: Cambridge University Press.
Frijda, N., & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In J. Russell & J. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 78–102). Paris: Cambridge University Press.
Garner. W. (1970). The stimulus in information processing. *American Psychologist, 25,* 350–358.
Garner, W. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
Izard, C. (1992). Basic emotions, relations among emotions, and emotion–cognition relations. *Psychological Review, 99,* 561–565.
Izard, C. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*(2), 288–299.
Jameson, K. A. (1996). Numerical scaling techniques for evaluating generative models of orthographies. *Acta Psychologica, 92,* 169–208.
Jameson, K. A., & Alvarado, N. (in Press). Color lexicons and cognitive salience: A study of color naming in Vietnamese, bilingual English–Vietnamese, and English populations. *COLOR: Research and Application*.
Kaiser, S., & Wehrle, T. (2001). Facial expressions as indicator of appraisal processes. In K.R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal theories of emotions: Theories, methods, research* (pp. 285–300). New York: Oxford University Press.
Levy, R. (1984). Emotion, knowing, and culture. In R. Shweder & R. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 214–237). New York: Cambridge University Press.
Melara, R. (1992). The concept of perceptual similarity: From psychophysics to cognitive psychology. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 303–388). Amsterdam: North-Holland.
Miller, G. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101,* 343–352. (Original work published 1956)
Moore, C., Romney, A. K., Hsia, T.-L., & Rusch, C. (1999). The universality of the semantic structure of emotion terms: Methods for the study of inter- and intra-cultural variability. *American Anthropologist, 101,* 529–546.
Nummenmaa, T. (1964). The language of the face. In *Jyvaskyla studies in education, psychology, and social research* (Vol. 9). Jyvaskyla, Finland: Jyvaskylan Yliopistoyhdistys, 66.
Ortony, T., & Turner, T. (1990). What's basic about basic emotions? *Psychological Review, 97,* 315–331.
Osgood, C. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology, 7,* 1–30.
Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press.
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.
Romney, A. K., Weller, S., & Batchelder, W. H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist, 88,* 313–338.
Rosenberg, E., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion, 19,* 111–138.
Russell, J. (1991). The contempt expression and the relativity thesis. *Motivation and Emotion, 15,* 149–168.
Russell, J. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion, 17,* 41–51.

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. (1997). Readings emotions from and into faces: Resurrecting a dimensional–contextual perspective. In J. Russell & J. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 295–320). Paris: Cambridge University Press.
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., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. *Journal of Personality and Social Psychology*, *67*, 186–205.
Russell, J., & Fernandez-Dols, J. (1997). What does a facial expression mean? In J. Russell & J. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 3–30). Paris: Cambridge University Press.
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.
Scherer, K. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Hillsdale, NJ: Erlbaum.
Scherer, K. (1992). What does a facial expression express? In K. T. Strongman (Ed.), *International review of studies of emotion* (Vol. 2, pp. 139–165). New York: Wiley.
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.
Smith, C. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology*, *56*, 339–353.
Smith, L. (2001). How domain-general processes may create domain-specific biases. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 101–131). Cambridge, UK: Cambridge University Press.
Smith, C., & Scott, H. (1997). A componential approach to the meaning of facial expressions. In J. Russell & J. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 229–254). Paris: Cambridge University Press.
Soja, N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development*, *65*, 918–937.
Thurstone, L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273–286.
Turner, T., & Ortony, A. (1992). Basic emotions: Can conflicting criteria converge? *Psychological Review*, *99*, 566–571.
Weller, S., & Romney, A. (1988). *Systematic data collection*. Thousand Oaks, CA: Sage.
Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. New York: Cambridge University Press.
Woodworth, R. (1938). *Experimental psychology*. New York: Henry Holt.