Journal of Experimental Psychology:
Human Perception and Performance
1997, Vol. 23, No. 1, 213–226

0096-1523/97/$3.00

Featural Evaluation, Integration, and Judgment of Facial Affect

John W. Ellison and Dominic W. Massaro
University of California, Santa Cruz

The paradigm of the fuzzy logical model of perception (FLMP) is extended to the domain of perception and recognition of facial affect. Two experiments were performed using a highly realistic computer-generated face varying on 2 features of facial affect. Each experiment used the same expanded factorial design, with 5 levels of brow deflection crossed with 5 levels of mouth deflection, as well as their corresponding half-face conditions, for a total stimulus set of 35 faces. Experiment 1 used a 2-alternative, forced-choice paradigm (either happy or angry), whereas Experiment 2 used 9 rating steps from happy to angry. Results indicate that participants evaluated and integrated information from both features to perceive affective expressions. Both choice probabilities and ratings showed that the influence of 1 feature was greater to the extent that the other feature was ambiguous. The FLMP fit the judgments from both experiments significantly better than an additive model. Our results question previous claims of categorical and holistic perception of affect.

The ability of organisms to perceive and identify displays of arousal of conspecifics seems ubiquitous in vertebrates and even in some invertebrates (i.e., squid, octopi, etc.). The face, of course, displays an informative signature of an organism's emotional state that is processed efficiently by cohorts. Brown and Dooling (1993) presented to parakeets images of scrambled and normal parakeet faces as stimuli. An important affective facial feature that parakeets can discriminate is the size of the iris. Male parakeets constrict the iris as part of their courtship displays to females, whereas the iris of a calm male parakeet is fairly large. Brown and Dooling (1993, Experiment 3) showed that female birds can indeed discriminate stimuli that differ only in pupil size. More important, they also determined that individual facial features could signal significant biological information, such as sex, age, or emotional arousal. These features were discriminated more quickly, and hence were more "perceptually salient," than features that could not provide this information.

The processing of affect is particularly well developed in humans, who appear to be able to recognize and characterize facial expressions of emotional affect in other humans with great accuracy and consistency (Collier, 1985; Ekman, 1993; Ekman & Friesen, 1975; Ekman, Friesen, & Ellsworth, 1972). Scientific research dealing with facial affect in humans dates back to the pioneering work of Duchenne de Boulogne (1862/1990), who used electrical stimulation of facial muscles in his participants to explore the muscle groups involved in the production of expressions of affect. One of several interesting findings was that not all muscles involved in affective expressions can be brought under conscious control (see, e.g., Duchenne de Boulogne, 1862/1990, p. 72). Although the current research was concerned with the recognition and identification of emotional affective expressions, we first discuss the literature on the processing of facial identity. That is, we operate on the testable assumption that the identity of faces is derived from the features that compose them in the same manner that the expression of a face is computed from facial features. Although it is necessarily the case that the features for facial expression differ from those for facial identity, the processing involved in these two domains could be similar. We make a distinction between information and information processing. The term *information* refers to the characteristics that are used in processing, whereas the term *information processing* refers to how these characteristics are processed. We hypothesize that previous findings of dissociations between emotion and identity may have resulted from differences in information rather than differences in information processing. Furthermore, segregated processing of identity and emotion in the brain (e.g., Sergent, Ohta, MacDonald, & Zuck, 1994) might reflect only differences in information rather than differences in information processing. Thus, until proved otherwise, we operate on the premise that studies of facial identity and facial emotion are equally informative about how people process and categorize facial information.

Recent research on the recognition and identification of faces has focused on issues of holistic versus featural identification (Brown & Dooling, 1993; Tanaka & Farah, 1993), categorical versus continuous perception (Etcoff & Magee, 1992), and mathematical modeling of face perception (Huber & Lenz, 1993). The hypothesis that facial recognition is a holistic process was explored by Tanaka and Farah (1993), who found that individual facial features were recognized more easily when displayed as part of a whole face than

John W. Ellison and Dominic W. Massaro, Department of Psychology, University of California, Santa Cruz.

This research was supported partly by grants from the Public Health Service (PHS R01 DC 00236), the National Science Foundation (BNS 8812728), and the University of California, Santa Cruz. We thank Michael M. Cohen for help at all stages of this research and appreciate the comments of Helen Kadlec.

Correspondence concerning this article should be addressed to Dominic W. Massaro, Department of Psychology, Social Sciences II, University of California, Santa Cruz, California 95064.

when displayed in isolation. Although recognition of individual features of faces was facilitated by the context of the whole face in normal orientation, no facilitation was found with scrambled faces, inverted faces, or pictures of houses. These findings suggested to Tanaka and Farah that facial recognition is in some sense a holistic process, differing qualitatively from the recognition of other types of images.

Holistic processing is a loaded term that is easily criticized, and, fortunately, researchers have begun to clarify what holistic processing means. Farah, Tanaka, and Drain (1995) and Carey and Diamond (1994) articulated two different characterizations of holistic processing. The terms *holistic encoding* and *configural* encoding are used to describe these two viewpoints. In holistic encoding, the parts of the face are not analyzed or represented explicitly. Rather, the face is analyzed and represented as a whole. As articulated by Tanaka and Farah (1993), "the representation of a face used in face recognition is not composed of representations of the face's parts, but more as a whole face" (p. 226). This viewpoint suggests that parts of the face are not the atoms of face analysis or representation. This viewpoint is closest to the traditional use of holistic processing because it bears great similarity to a template matching scheme of pattern recognition (Massaro, in press). According to this viewpoint, the parts of the face would not be as accessible as the complete face. This type of model is consistent with Tanaka and Farah's finding that individual facial features were more easily recognized when they were part of the complete face than when they were presented alone.

The second characterization of holistic processing, *configural encoding,* refers to the possibility that the spatial relations among the parts of the face are more influential than the parts themselves. According to this view, it is not assumed that the parts are not represented but that it is simply the relations among the parts that are critical for analysis. This interpretation of holistic processing also is consistent with Tanaka and Farah's (1993) finding that individual facial features were more easily recognized when they were part of the complete face than when they were presented alone. In this case, the complete face would provide spatial relations that would not be available in a part of the face presented in isolation. Even with this clarification of holistic processing, it is difficult to distinguish these two types of "holistic" processing from the more general operation of context effects in perception, such as those that produce the so-called "word superiority effect" in letter recognition tasks (Massaro & Cohen, 1994). One of the goals of this research was to quantify and test models of the recognition of facial affect so that the field is not limited by the ambiguity of verbal theories.

Etcoff and Magee (1992) presented human observers with computer-averaged human faces that differed by constant increments along a dimension of emotional affect. Three different tasks used stimulus faces generated by weighted averaging of line drawings of exemplar faces displaying several emotional expressions: (a) an ABX task in which the first two stimuli were different and the participant identified which one was a match to the third stimulus; (b) an identification task in which participants were shown stimulus faces in random order and indicated to which affective category (i.e., sad or happy, etc.) the stimulus face belonged; and (c) a free-naming task in which participants responded with a short description of the displayed expression, along with a description of the situation that would produce such an expression. Given the similarity of their results to previous findings of "categorical perception," Etcoff and Magee concluded that all facial expressions except surprise were perceived categorically. However, because their discrimination tasks encouraged participants to categorize rather than discriminate (Massaro, 1987a), the issue of categorical perception was not really addressed by this study in spite of the researchers' interpretations.

It is now well-known that discrimination tasks do not necessarily measure discrimination capacity (Massaro, 1987a). Many discrimination tasks have memory limitations and performance is easily influenced by category labels. The ABX task, for example, makes it difficult to compare the third stimulus X with the first stimulus A. In this task, participants often encode the stimuli categorically and base their discrimination decision on these category labels. Thus, better discrimination between items from a different category than items from the same category does not conclusively show categorical perception.

Given a stimulus continuum between two alternatives, a typical result is that the identification judgments are an ogival function of changes along the stimulus continuum. Several researchers, such as Etcoff and Magee (1992), have interpreted these prototypical findings as evidence for categorical perception. These investigators have concluded that pairs of equally spaced stimuli are perceived discontinuously. Two stimuli within a category are supposedly more poorly discriminated than two stimuli from two different categories. The error in this interpretation, however, is that the dependent measure, the proportion of judgments, is being treated as a linear measure of perception. However, it has been shown that this type of result follows directly from continuous perception (Massaro, 1987a, 1987b). Sharp identification boundaries follow naturally from a system with continuous information and a decision criterion. In the current research, for example, we found that perceivers had continuous information about each of the two facial characteristics being varied. Even though the results are best described by a model with continuous information, the results show the same "discontinuous" identification functions that have been interpreted previously in favor of categorical perception. The lesson from this exercise is that the shape of the identification functions alone is not sufficient to conclude whether perception is continuous or categorical. The most direct measure involves the quantitative tests of mathematical models that assume either continuous or categorical information (Massaro, in press). Finally, in the Etcoff and Magee (1992) study, the results for the emotion surprise did not give the prototypical results. Although Etcoff and Magee could not account easily for this discrepancy, it is explained more easily when the hypothesis of categorical perception is rejected for all affective dimensions.

Researchers have used mathematical modeling techniques for the analysis of facial feature perception. For example, in a recent study, Huber and Lenz (1993) tested a feature learning theory of animal concept discrimination. They trained pigeons to discriminate among two sets of schematic human faces that differed along four feature dimensions of three values each. The three feature values were arbitrarily set at +1, 0, and −1. Thus, the sum of the feature values in a given schematic face could range from +4 to −4. The pigeons were able to learn to discriminate reliably faces with positive feature sums from those with negative feature sums using a successive go/no-go procedure with food pellets as the reinforcer and peck rate as the dependent variable.

Although Huber and Lenz (1993) did not address specifically the issue of affect recognition, their stimulus features were varied in a way that roughly corresponded to variations in affective features in actual human faces. That is, eyebrows were raised and lowered, mouth-to-chin distances and eye separation were varied, and nose appearance was varied. Huber and Lenz concluded that their results could be explained by a categorization according to a sum of the feature values (i.e., an additive combination). They were able to predict the classification decisions of their pigeons fairly well using this linear additive feature model. However, they did not test these results against other models that assume some other combination of feature values in pattern recognition.

The aforementioned studies all incorporated some kind of pattern-recognition task involving continuous or quasi-continuous stimulus input and generating more or less arbitrarily categorical responses. The *fuzzy logical model of perception (FLMP)* (Massaro & Cohen, 1990, 1993; Massaro & Ferguson, 1993) has been shown to provide superior predictive capability in similar experiments in the realm of speech perception and other domains and also could lend itself to analysis of facial affect recognition and categorization paradigms. Because analyses under the FLMP also can allow participants to respond to stimuli with a continuum of responses rather than an arbitrarily forced categorization, the model should allow a more precise exploration of the question of the processing of a continuum of facial affect changes (Massaro, 1989, p. 41). In addition, the FLMP is able to model two-choice identification responses of participants who are required to respond to several levels along a stimulus dimension.

Previous research on affect recognition and classification generally has used either highly schematic (i.e., line drawing) stimuli or exemplar photographs of people who are either naturally displaying emotional expressions or feigning (as an actor might) the expression in question. A third method (Ekman, 1993; Ekman et al., 1972; Ekman, Hager, & Friesen, 1981) has used photographs of people who are instructed to hold a specific facial pose (i.e., lips held a certain way, brows deflected downward, etc.); these unaltered photographs are used or can be dissected and recombined in various composites. These types of stimuli are inherently limited in several ways. First, drawings or pictures of actual human faces always will be confounded by familiarity, attractiveness, covariation of features, and so on, which make the stimuli virtually impossible to standardize. Second, research has shown that feigned expressions may not be identical or even comparable to genuine affective expressions (Duchenne de Boulogne, 1862/1990; Ekman et al., 1981). Third, most previous research, with the notable exception of the work of Huber and Lenz (1993), has used stimulus continua formed by averaging or "morphing" across the entire face, thus obscuring the relative contributions of individual features to the perception of affect. Fourth, the identification of affect in faces in which only incomplete, ambiguous, or contradictory information is available has not been explored because of the limitations of stimuli and constraints of research paradigms.

Responses to incomplete, ambiguous, or contradictory stimuli are potentially the most challenging to information-processing models of cognitive processes and could be used to distinguish between rival models of facial affect perception (Massaro, in press). The goal in the current research was to test among alternative models in this situation. An important assumption of models is whether perception of a stimulus feature is assumed to be categorical or continuous. Categorical and continuous models make different predictions about affect identification and rating; these differences can only be teased apart using factorial designs involving multiple levels of several independently varied features. Expansion of the factorial designs to include cases in which information from a particular feature is presented in isolation can provide further tests of competing models. Holistic encoding would seem to claim that facial identifications and ratings of a complete face could not be predicted from the identifications and ratings of parts of the face presented in isolation.

These considerations prompted us to seek a set of stimuli for facial affect research that would be standardized and replicable, as well as controllable, over a wide range of feature dimensions. Research in the Perceptual Science Laboratory at the University of California, Santa Cruz, has developed a computer system capable of displaying a highly realistic but fully controllable synthetic talking face (Cohen & Massaro, 1993, 1994). This technology was used in the current research to overcome some of the limitations of the facial stimuli used in previous studies. Although the face is realistic, its parts are independently controllable, fully quantifiable, and easily replicable. The primary advantage of this synthetic face is that displays of ambiguous or contradictory feature deflections, or partial face presentations, can be made more easily than with previous types of facial stimuli (see Figures 1 and 3). In addition, we used the expanded factorial design illustrated in Figure 1. The advantages of the expanded factorial design are that single features as well as all feature combinations are tested. This design provides a stronger test of models of perceptual recognition and judgment (Massaro & Cohen, 1990).

Accordingly, we designed two experiments to address important issues in the field of face processing. First, to what extent can the identification and rating of a face be described by the processing of the parts that make it up? Second, is a part or dimension of a face perceived contin-

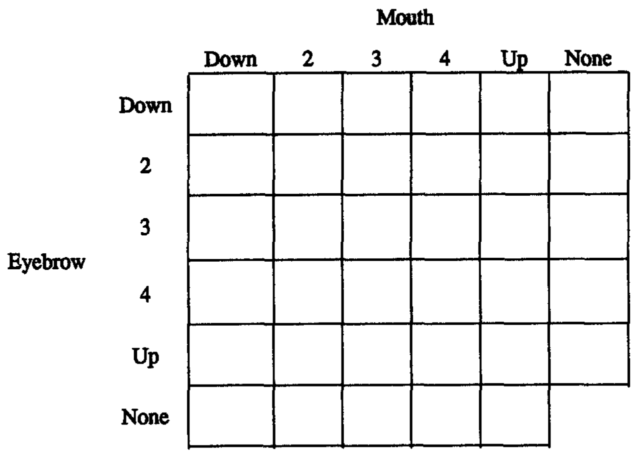

Figure 1. Expansion of a typical factorial design to include upper face and lower face conditions presented alone. The five levels along the eyebrow and mouth continua represent deflections of the displayed feature along five physical steps from downward displacement to upward displacement.

uously or categorically? Third, are separate parts of the face evaluated independently of one another? Fourth, how are the parts of the face combined or integrated to achieve facial recognition? These important questions are generated and addressed by extending our paradigm of inquiry (Massaro, 1987b) to the exploration of facial affect recognition using computer-generated faces displaying features corresponding to typical human emotional expressions. If the FLMP does indeed describe the processes involved in facial affect recognition, it should provide more accurate predictions than those provided by holistic models, additive linear feature models (Huber & Lenz, 1993), or categorical models (Etcoff & Magee, 1992).

The FLMP

The FLMP assumes three stages or operations between some stimulus event and a response (see Figure 2). Specifically, with regard to the current experiments, the FLMP predicts that participants evaluate expressions of emotional affect according to information arriving from multiple sources (i.e., individual facial features). Stimulus variation is made along several dimensions (in this case, two: brow deflection [BD] and mouth corner deflection [MD]). The information from each source is evaluated according to the degree of match to a prototype or degree of support for a particular affect (happy vs. angry). This information then is integrated according to a multiplicative formula to determine how representative the stimulus is of a particular affective class. The decision process determines the relative goodness of match of the stimulus with each prototype and the appropriate response is made.

Several issues had to be resolved to justify our paradigm of investigation. One had to do with the naturalness of our synthetic face and the extent to which results with this face could be generalized to the processing of real faces. A second concerned the emotional categories that were used and the features that were varied to produce these emotions. As described in the introduction, the synthetic face offers an ideal compromise between naturalness and control. The face was created to mimic all the characteristics of a real face while permitting the exact control of its different parts. A reader might claim that there is no reason to believe that participants actually perceived the faces as happy or angry. However, no feedback was given in our experiment, so the completely consistent and orderly results for each of the 48 participants falsifies this claim. Both feature dimensions influenced judgments in the appropriate direction. Although we did not build in all possible cues, the synthetic face clearly had characteristics that corresponded to these emotions on normal faces. The goal was not to create faces with all possible cues but to address the question of how multiple cues are evaluated and integrated in the perception of facial affect.

In the current research, we chose the affective categories *happy* and *angry* because they represent two of the basic categories of emotion. We asked participants to judge the more general affective categories of happy and angry rather than specific facial expressions such as smile and frown. The latter task would tend to direct attention to specific facial characteristics, whereas our goal was to judge the whole face. We do not view happy and angry as end points on a single affective dimension. Even so, we believe we can create faces that vary in the degree to which they represent one emotion as opposed to others. For practical reasons, we limited our research to the categories of happy and angry. Finally, we had to choose which features to vary systematically to create a range of emotions between happy and

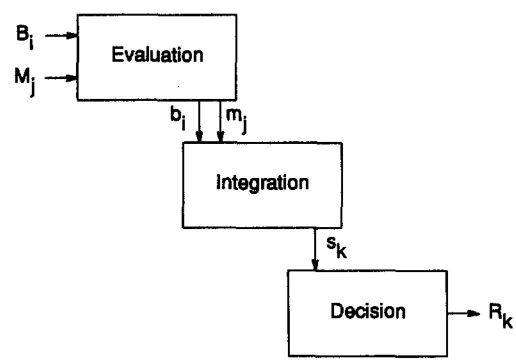

Figure 2. Schematic representation of the three stages involved in perceptual recognition. The three stages are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. The sources of information are represented by uppercase letters. Eyebrow deflection is represented by B_i and mouth deflection by M_j. The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters b_i and m_j). These sources then are integrated to give an overall degree of support, s_k, for a given affect alternative k. The decision operation maps the outputs of integration into some response alternative, R_k. The response can take the form of a discrete decision or a rating of the degree to which the alternative is likely.

angry. We believe that emotion categories are fuzzy in that no set of necessary and sufficient features characterizes a particular emotion. We chose two features that seem to differ somewhat in happy and angry faces. Again, not all happy and all angry faces will differ on these features, but the features are correlated with the emotion categories. Even for natural faces, there is some controversy concerning the degree to which observers can accurately categorize different emotions. As concluded by Fridlund (1994, p. 237), there is no evidence for the claim that a given facial expression is linked unambiguously with a single emotion category.

We realize that many other features are correlated with these affective categories. For example, there is a tendency for a tightening around the eyes and a lifting of the cheeks in spontaneous smiling (Allen & Atkinson, 1981; Duchenne de Boulogne, 1862/1990; Ekman et al., 1981). Although we can achieve this pose in our animated face, we had to limit our study to just two features to keep the number of unique faces reasonably small and the number of test observations relatively large. Our test faces might be most comparable to Ekman's (1973) expression type called *facial emblems*, which appear only in the context of social interactions.

An important criterion for manipulating two features is that they can be varied independently of one another. Thus, varying one cue in the upper face and one cue in the lower face was an ideal solution. Furthermore, there appear to be upper motor neurons from the neocortical motor strip in which the upper and lower face are served by different neurons (Fridlund, 1994, pp. 92–94). In the current experiments, five levels of the upper face and five levels of the lower face were combined factorially. In addition, the 10 half-face conditions were presented, as prescribed by an expanded factorial design. The features varied were BD and MD, as can be seen in Figure 3. BD was varied from somewhat elevated and arched for a prototypically happy affect to fully depressed and flattened for a prototypically angry affect. MD was varied from fully curled up at corners for a prototypically happy affect to fully curled down at corners for a prototypically angry affect. The maximum feature deflections were obtained by comparison to features displayed in exemplar photographs in Ekman and Friesen (1975).

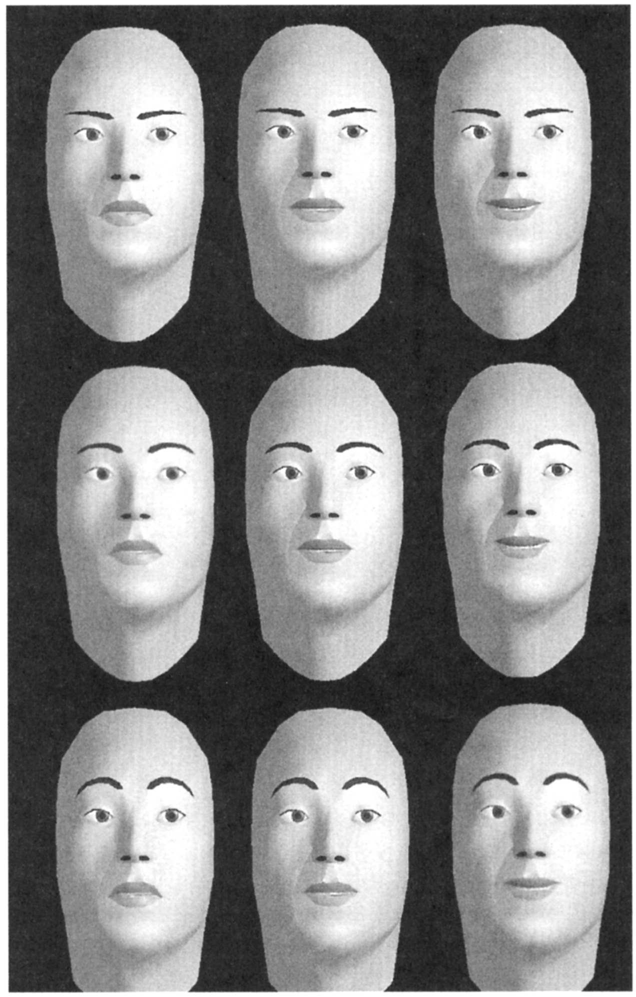

Figure 3. The four faces displaying the maximum feature deflections (at the corners of the figure) as well as faces displaying "neutral" deflections. The center face is the "neutral baseline" face, with both mouth and brow deflected at the neutral values. Note that some faces are ambiguous and incongruent in their expressions. The unimodal (half-face) conditions displayed only the upper or lower half of the stimulus face. The reader can cover half of the face to experience these conditions.

Within the context of the FLMP, it is assumed that participants generate prototypes corresponding to happy and angry affects. The prototype corresponding to a happy face might consist of the following description: The eyebrows are somewhat elevated and arched, and the mouth corners fully curled up. The prototype corresponding to an angry face, on the other hand, might consist of the following description: The eyebrows are fully depressed and flattened, and the mouth corners fully curled down.

All other sources of information contributing to facial affect are not listed in the prototype descriptions because it is assumed that they are not being influenced systematically by the independent variables BD and MD. At the feature evaluation stage, each physical input is transformed to a psychological value and is represented in the model equations in lowercase (e.g., if B_i represents the brow information, B_i would be transformed to b_i, the degree to which BD supports the alternative happy, H). An important assumption is that the evaluation of a particular feature occurs independently of the presence or absence of other features and their information value. The evaluation of a specific feature in the upper half of the face would produce the same result when presented in the context of the complete face and when presented without the lower half. With just two alternatives, H and angry (A), we can make the simplifying assumption that the degree to which BD supports the alternative A is $1 - b_i$ (Massaro & Friedman, 1990). Feature evaluation occurs analogously for the feature mouth deflection, M_j. Feature integration consists of a multiplicative combination of the feature values supporting a given alternative. If b_i and m_j are the values supporting alternative H, then the total support, $M(H)$, for the alternative H would be given by the product of b_i and m_j:

$$M(H) = b_i m_j. \tag{1}$$

The third operation is decision, which uses a relative goodness rule (RGR; Massaro & Friedman, 1990) to give the relative degree of support for each of the test alternatives. In the two-alternative choice task, the probability of a happy choice, $P(H)$, is equal to

$$P(H|B_i, M_j) = \frac{M(H)}{M(H) + M(A)}, \tag{2}$$

where $P(H|B_i, M_j)$ is the predicted choice given stimulus B_i, M_j. The predictions for rating judgments have the same form as those for choice judgments. Thus, the rating of the degree of happiness given stimulus B_i, M_j is

$$R(H|B_i, M_j) = \frac{M(H)}{M(H) + M(A)}, \tag{3}$$

where $R(H|B_i, M_j)$ is the predicted rating. The actual ratings between 1 and 9 are transformed linearly into values between 0 and 1 by the transformation

$$R(0\text{–}1) = \frac{R(1\text{–}9) - 1}{8}, \tag{4}$$

where $R(0\text{–}1)$ is the transformed rating and $R(1\text{–}9)$ is the actual rating. This transformation is based on the assumption that observers are able to rate the test items along an interval scale on the continuum of interest. The FLMP requires five free parameters for the levels of BD and five for the levels of MD.

Additive Model of Perception

The FLMP assumes a multiplicative combination of feature values representing the different sources of information. Other types of models have been proposed to explain performance in various other domains of pattern recognition (Cohen & Massaro, 1992; Cutting, Bruno, Brady, & Moore, 1992; Massaro, 1988; Massaro & Cohen, 1993). For example, in an additive model (cf. Huber & Lenz, 1993), it is assumed that the sources of information available to the participant are added rather than multiplied as in the FLMP. One instantiation of this model is exactly the same as the FLMP except that the feature values are added rather than multiplied. Adding the values at integration with an RGR at decision reduces to an averaging model (Massaro, 1987b, chap. 7). In addition, this model can be made more general by allowing one featural dimension to have more influence than the other. Predicted ratings of this weighted averaging model (the additive model of perception [AMP]) are given by

$$R(H|B_i, M_j) = wb_i + (1 - w)m_j, \tag{5}$$

where w is the weight given to BD and $(1 - w)$ is the weight given to MD.

The AMP requires five free parameters for the levels of BD and five for the levels of MD, as in the FLMP. An additional free parameter also is necessary to accommodate the weight term. Note that the AMP is mathematically equivalent to a single-channel model in which the participant attends to information from just one modality or feature on a particular trial.

Holistic and Categorical Models

Unfortunately, both holistic and categorical models are not easily formalized to make testable predictions for this task. For both types of models, it might be claimed that perception of each face is unique and cannot be predicted from performance on the parts that make it up. For holistic processing, the processing of the complete face cannot be reduced to processing of the separate parts. Holistic processing contrasts sharply with the FLMP's assumption that evaluation of the features occurs independently of one another. Similarly, a categorical model might predict that categorization of the face is uniquely determined by its complete configuration. The categorical model also claims that affect perception is discrete in that gradations of affect are not easily perceived within an affect category. However, both viewpoints would have to claim that the FLMP should fail because its assumptions are the antitheses of these models. Thus, we take an adequate description by the FLMP as evidence against these models. Furthermore, the reaction times (RTs) of the perceptual judgments provide an additional test of categorical perception. If perception of the complete face is categorical, then the RTs should be independent of the ambiguity of the facial display. It also is possible that the separate features of the face are perceived categorically. There is a model that is mathematically equivalent to the AMP that allows us to provide a test of this specific version of categorical perception. In this model, information from each feature is perceived discretely, and the response is generated from this discrete information.

General Method

Because both experiments used the same apparatus, stimulus set, and subject pool, we describe them together. We then describe the procedures used in each experiment, along with reports of the results. Next we present a section describing the tests of the alternative models being considered.

Participants

Forty-eight students from the undergraduate psychology subject pool at the University of California, Santa Cruz, were involved in the current experiments. The participants were involved in the subject pool as part of their required undergraduate psychology coursework. Because other researchers have found that men and women may perceive affect differently (Harrison, Gorelczenko, & Cook, 1990; Kirouac & Dore, 1985, 1983), we tested participants of both sexes. The students ranged in age from 18 to 40 years (M = 22.4 years), and they had normal visual acuity. We originally had intended to include comparisons of performance of participants with different hand (and presumably cerebral) dominance, but of the 48 participants, we found only 4 left-handed individuals, and this was an insufficient sample for valid comparison.

Apparatus

The stimuli used in these experiments were generated by facial synthesis software using a parametrically controlled polygon topology synthesis technique with texture-mapped skin surfaces and ray-tracing lighting simulation. This program, Face39, is capable of producing animated visible speech at 60 frames per second synchronized with the output of an auditory text-to-speech synthesizer, although these capabilities were not used in the current experiments. A complete description of the technology was given by Cohen and Massaro (1993, 1994).

A set of Face39 stimuli was constructed to portray affective expressions that varied along two feature dimensions in an expanded factorial design. The features varied in the stimulus set were BD and MD. These two features were varied in the two-factor condition along continua from maximum down deflection to maximum up deflection in five steps; one-factor (i.e., half face; brows only [UB] or mouth only [UM]) conditions also were included at the same five levels. This stimulus set, then, incorporated 35 faces: 25 in the two-factor condition and 5 in each one-factor condition (see Figure 1).

These features were chosen to approximate expressions of anger (at the maximum down value of both features) and happiness (at the maximum up value of both features) by comparison with exemplar faces from Ekman and Friesen (1975). The Face39 expressions are only approximate, however, because of the general lack of correspondence between the wire-frame polygon structure of the face model with the actual muscular articulation of the human face. Another constraint limiting the realism of the synthetic expressions is the requirement, in the current experiments, to vary only two mutually exclusive features within the stimulus faces. Actual expressions of affect in humans usually involve covariation in multiple features (Duchenne de Boulogne, 1862/1990; Ekman et al., 1972) and so are more difficult to implement realistically on the wire-frame polygon model.

Although previous studies of categorization or rating of affect have used stimuli formed by varying the entire face along continua between two end-point expressions (cf. Etcoff & Magee, 1992), the present expanded factorial design varied the two features BD and MD independently of one another. This design requires the presentation of faces with inconsistent as well as consistent features. The experiments included some stimulus faces that were ambiguous and that did not strictly correspond to either of the response alternatives (see Figures 1 and 3). In addition, the expanded factorial design presents each of the features in isolation. An important property of this design is that it requires models to predict both the single feature and complete face conditions together.

The Face39 program and the experimental control programs used to run the experiments and collect participant data were implemented on a Silicon Graphics 4D/Crimson VGX workstation running under the IRIX operating system; the stimulus faces were displayed to the participants on 12-in. (30.48 cm) NEC Model C12-202A color monitors; and participant responses were collected on TVI 950 video display terminals (VDTs) and their associated keyboards. This system is capable of recording RTs with millisecond accuracy (Cohen & Massaro, 1994). Data analysis was performed on the same Silicon Graphics workstation using FORTRAN 77 data analysis routines and on a Sun workstation using the SAS statistical package (SAS Institute, Cary, NC).

Experiment 1

Method

Twenty-six participants were tested in a two-alternative, forced-choice (2-AFC) task. Participants were required to respond to each stimulus face with either "happy" or "angry" by pressing a correspondingly labeled key on either the left or right edge of the VDT keyboard. No intermediate responses were allowed. The order of the responses on the keyboards was counterbalanced across subjects, with 12 participants seeing the order happy–angry and 14 participants seeing angry–happy.

After a short title sequence, the control program began displaying the individual face stimuli on the monitors in the participants' cubicles. The faces were displayed for 1,000 ms each, with a 100-ms, 1000-Hz orienting beep being played 700 ms before stimulus onset. The stimulus faces (cf. Figure 3) were sized to fill the vertical dimension of the 12-in. (30.48-cm) monitor screens (15 cm high) and were viewed at a distance of about 45 cm. No visual fixation point was provided. Ratings were collected, and RT latency was recorded from stimulus onset. Because the control program collected all 4 participants' responses before displaying the next face in the stimulus set, there was a short but variable time between trials, on the order of 3–4 s.

Each experimental session included 10 practice trials and 280 stimulus trials; the stimulus trials were selected from the stimulus set according to a random selection without a replacement protocol, which resulted in each stimulus face being displayed eight times per session (not including 10 practice trials). Each participant was involved in two experimental sessions, separated by a 5-min rest period, and so saw each stimulus face 16 times.

Results

Because the participants in this experiment were limited to two choices, their mean responses to a particular stimulus face could be expressed as a probability of identifying the face as being happy ([*P*]happy). The probability of an angry identification was, of course, 1 − (*P*)happy. Figure 4 shows the average results in this 2-AFC task. A concern for other researchers might be the questions of whether the faces were ecologically valid and whether the participants actually perceived the faces as happy or angry. These questions appear to be favorably answered because Figure 4 demonstrates conclusively that all 26 participants were perfectly accurate on the most unambiguous faces. The face with both the mouth and brow deflected downward the most was always categorized as angry, and the face with both the mouth and brow deflected upward the most was always categorized as happy. This unanimous behavior is particularly impressive because the participants were never given any instruction or feedback concerning the intended meaning of the features. This behavior provides evidence that the features we manipulated were meaningful and realistic and that the results are externally valid.

As expected, identification performance varied systematically with changes in the independent variables. The BD and MD features influenced performance in the predicted manner. The upper left face in Figure 3 was unanimously identified as angry, whereas the lower right face was always called happy. Figure 4 shows that the probability of a happy

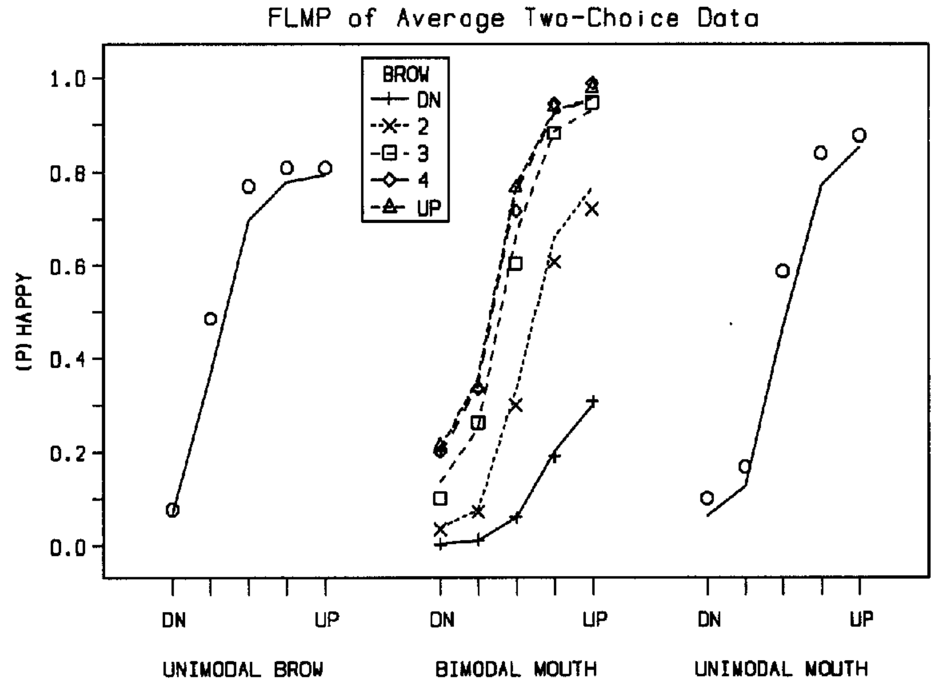

Figure 4. Predicted (lines) and observed (points; *P*[happy]) identification judgments in Experiment 1 as a function of the brow and mouth conditions. Brow DN corresponds to the eyebrows fully depressed and brow UP to the eyebrows somewhat elevated and arched. Mouth DN corresponds to the mouth corners fully curled down and mouth UP to the mouth corners fully curled up. Predictions are for the fuzzy logical model of perception (FLMP).

identification increased as the brow changed from a fully depressed and flattened position to a fully elevated and arched position. Analogously, the probability of a happy identification increased as the mouth changed from being fully curled down at the corners to being fully curled up. There also was a statistically significant interaction in that the influence of one variable was larger to the extent the other variable was neutral or ambiguous (all *F*s were significant at the .001 level).

RTs of the identification judgments also were analyzed. A subset of the 35 stimulus events was partitioned into happy, neutral, and angry classes, and a mean RT was computed for each participant for each class. The happy faces were those that had both features at Level 4 or 5 and the single-feature conditions at either Level 4 or 5. Similarly, the angry faces were those that had both features at Level 1 or 2 and the single-feature conditions at either Level 1 or 2. The neutral faces were the single-feature conditions at Level 3 and the two-factor conditions with one feature at Level 3 and the other feature at either Level 2, 3, or 4. No significant difference was found between RTs to angry and happy expressions, but these RTs were significantly faster than RTs to neutral expressions: mean RT for happy = 1,087 ms (SD = 317 ms), angry = 1,084 ms (SD = 321 ms), and neutral = 1,200 ms (SD = 322 ms), $F(2, 48) = 12.251$, $p <$ .001. No significant interaction was found between participants' gender and the affective expression of the stimulus face.

Experiment 2

Method

Twenty-two participants were tested in the rating task. Participants received instructions to rate the affect they perceived on the stimulus face using a scale from 1 to 9 on the keyboard. This scale was counterbalanced across subjects for order of affect; 12 participants encountered the order angry–happy (1–9) and 10 participants encountered the order happy–angry (1–9). The participants were told to rate faces that clearly exhibited the affect in question as a 1 or a 9, faces that were perfectly ambiguous between the affects were to be rated 5, and intermediate faces were to be rated using the 2, 3, 4 keys (or 6, 7, 8) depending on the degree of perceived correspondence to the happy or angry affects. Participants were not shown any exemplar faces, nor were they given any feedback during the rating task. Participants saw 10 practice trials and 280 stimulus trials per session, as in Experiment 1. There were two sessions separated by a 5-min break. Each stimulus face was therefore rated 16 times by each participant.

Results

The individual participant's 16 ratings for each stimulus face were averaged, yielding 35 data points per participant. Figure 5 shows the ratings averaged across the participants. As in the two-choice task, the independent variables influenced performance in the predicted manner. The average rating of happy increased as the eyebrows became more

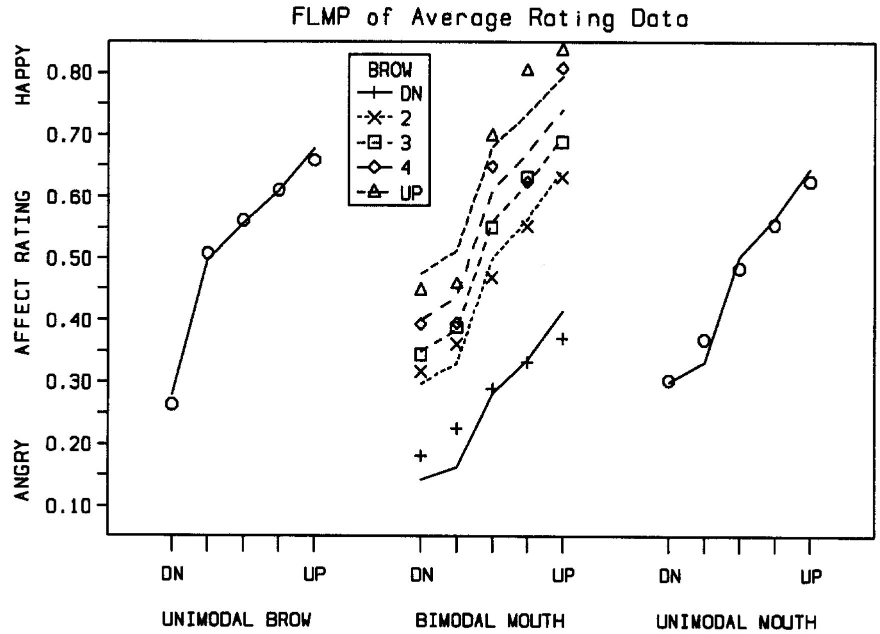

Figure 5. Predicted (lines) and observed (points; *R*[happy]) rating judgments in Experiment 2 as a function of the brow and mouth conditions. Brow DN corresponds to the eyebrows fully depressed and brow UP to the eyebrows somewhat elevated and arched. Mouth DN corresponds to the mouth corners fully curled down and mouth UP to the mouth corners fully curled up. Predictions are for the fuzzy logical model of perception (FLMP).

elevated and arched and as the corners of the mouth curled up. There also was a statistically significant interaction in that the influence of one variable was larger to the extent the other variable was neutral or ambiguous (all *F*s were significant at the .001 level).

Test of the Models

The FLMP makes specific predictions about how participants should perform when viewing stimulus faces involving features that are independent of one another. The probability of identifying or rating a particular stimulus face as either happy or angry should depend on continuous information from both features. Identification probabilities and ratings should be more extreme when features are congruent and unambiguous, whereas they should be less extreme when features disagree or are neutral or ambiguous. Furthermore, as the ambiguity of one feature increases, judgments should be influenced more by the other, less ambiguous feature. To the degree that the outcome of feature integration is ambiguous, longer RTs of these judgments should be observed.

Model fitting was accomplished using the STEPIT subroutine (Chandler, 1969), which finds local minima of real functions in several parameters. In our tests, the subroutine minimized the root mean square deviation (RMSD) between an individual participant's observed data points and a set of data points predicted by the 10 free parameters (11 in the case of the AMP) of a multivariate regression equation relating the stimulus feature values to the observed data. The subroutine iterates the model equations, changing parameter values until values are found that minimize the RMSD between the observed data and the predicted data. The value of the final RMSD is an indication of the model's goodness of fit. The models are tested against the results of individual participants, allowing an exploration of individual differences.

Given the 5 × 5 expanded factorial design, 10 free parameters are necessary to fit the FLMP to the 35 data points: five parameters for each level of BD and MD. The parameters represent the degree to which these features match those in the happy prototype. The FLMP and AMP were fit to each of the individual participants and to the mean participant computed by averaging the results across participants. As can be seen in Figure 4, the predictions of the FLMP did reasonably well in capturing the trends in the data. In the two-choice identification task, the FLMP's RMSDs for the participants ranged from .0465 to .1277, with an average RMSD of .0821. The fit of the mean participant gave an RMSD of .0415. The fit of the AMP produced larger RMSDs, with a range between .0902 and .2100, and an average RMSD of .1547. The fit of the AMP to the mean participant data was .1160. Figure 6 shows the observed results along with the predictions of the AMP. The FLMP provided the best fit to the individual participants as well as to the average results.

A comparison by participant of the RMSD fits of the two models is informative. Figure 7 gives the RMSD values for

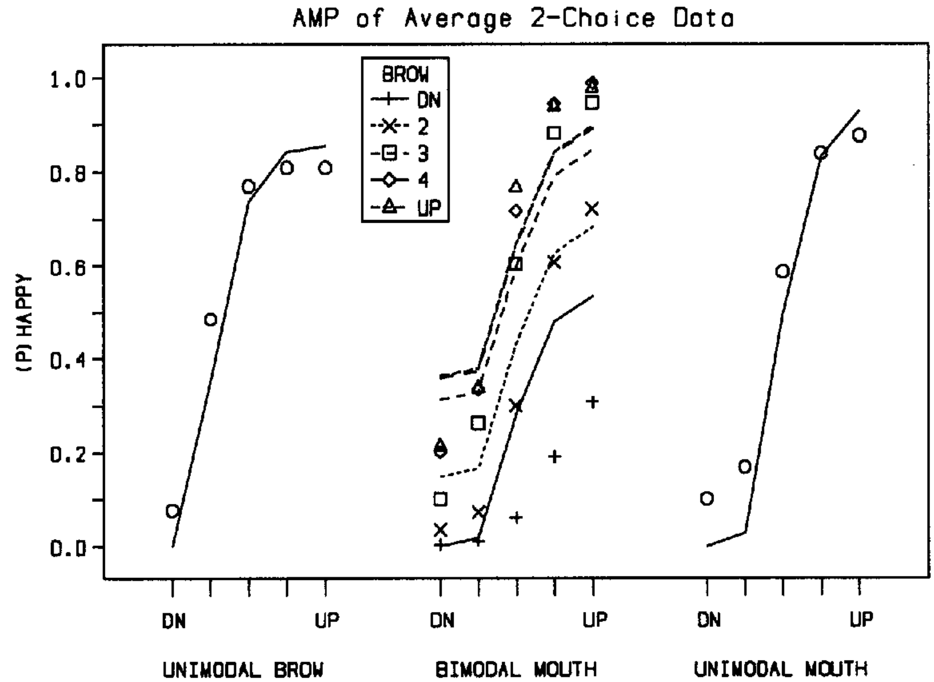

Figure 6. Predicted (lines) and observed points (*P*[happy]) identification judgments in Experiment 1 as a function of the brow and mouth conditions. Brow DN corresponds to the eyebrows fully depressed and brow UP to the eyebrows somewhat elevated and arched. Mouth DN corresponds to the mouth corners fully curled down and mouth UP to the mouth corners fully curled up. Predictions are for the additive model of perception (AMP).

each participant for the fit of the FLMP and AMP. Although data from 2 participants were fit better by the AMP, the other 24 participants showed better fits by the FLMP (see Figure 7). The RMSD fits of the two models were compared, and the FLMP provided a significantly better overall fit than did the AMP, $F(1, 25) = 69.197$, $p < .001$.

The poor fit of the AMP relative to that of the FLMP also provides evidence against the discrete perception of each of the two stimulus features. In this model, which is mathematically equivalent to the AMP, the discrete information from each of the two features is used to make the identification judgment (Massaro, in press). Of course, other categorical models are possible, and one of these might provide an adequate description of the results. Obviously, one cannot reject all possible categorical models, but there are no other known quantitative formulations of this theoretical notion. Another investigator is always free to develop another version and test it against our results.

Unfortunately, we know of no holistic model that can be tested quantitatively against the results. One class of holistic models would assume that each unique feature combination would create a unique affect that could not be predicted from some simple combination of the two features. This formulation would capture the essence of holistic models that somehow the whole is more than some combination of its parts. We could not test a specific quantitative formulation of this holistic model because it requires as many free parameters as observed data points. Therefore, this model is untestable. However, the excellent fit of the FLMP, which assumes independence at the feature evaluation stage, provides evidence against this type of holistic model. If each combination of features is unique, then a model assuming independence between features should fail. The fact that the FLMP does not fail therefore is evidence against holistic perception.

Other types of holistic models are not so easily falsified. It certainly is possible that the features being used are better described in terms of spatial relations among parts of the face rather than only the features. The use of spatial relations must be true at some level. For example, the deflection of the corner of the mouth is probably evaluated relative to the center of the mouth. The deflection of the eyebrows could be evaluated relative to the eyes and nose. Although the present research cannot address this issue, our research paradigm is rich enough to investigate what is actually being used by the perceiver. The deflection of the eyebrows could be varied orthogonally to the distance between the eyebrows and eyes. Experimental manipulations of this nature would allow the investigator to zero in on the actual information being used.

Although the FLMP provides a significantly better fit than the AMP, it is valuable to determine how good the fit is in an absolute sense. A benchmark measure has been developed to provide this index of goodness of fit of a

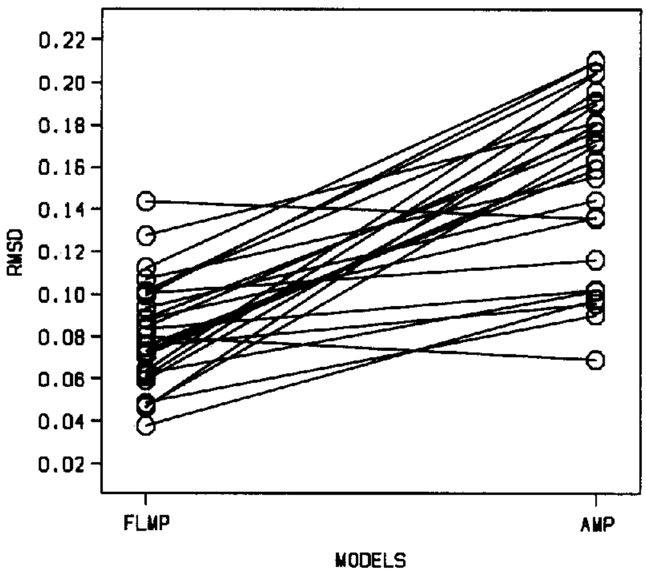

Figure 7. The root mean square deviation (RMSD) values for each participant for the fit of the fuzzy logical model of perception (FLMP) and the additive model of perception (AMP).

model (Massaro & Cohen, 1993). Even if a model is perfectly correct, we cannot expect it to fit results perfectly. The FLMP is deterministic (has no variability) at the feature evaluation and integration processes and becomes probabilistic at the decision process. The variability at the decision process is attributable to the RGR in which the probability of a response is equal to the merit of that alternative relative to the sum of the merits of all relevant alternatives. For example, given an RGR value of .8, that alternative is chosen .8 of the time. With a finite number of observations, we cannot expect that the actual proportion of responses with the alternative will be precisely .8. Thus, we can expect some mismatch between the predicted and observed values, even if the model is correct. The observed variability should be equal to that expected from probability theory—in this case, on the basis of simple binomial variability. It is possible to determine the expected binomial variability as a function of the observed response probabilities and the number of observations of each experimental condition. With this prediction of the expected variability, we can ask whether the fit of a model is poorer than what would be expected from chance binomial variability. The standard deviation of the mean of a binomial distribution (with two outcomes) is equal to the square root of its binomial variance:

$$\sigma = \sqrt{\frac{pq}{N}} \tag{6}$$

where p is the probability of one outcome, q the probability of the other ($q = 1 - p$), and N is the number of observations.

The benchmark RMSD is determined by computing the binomial variance for each of the 35 experimental conditions, averaging these 35 values, and taking the square root:

$$\text{RMSD(b)} = \sqrt{\frac{\sum_{1}^{k} (pq/N)}{k}}, \tag{7}$$

where RMSD(b) is defined as the benchmark RMSD. These RMSD(b) values can be compared with the RMSD values from the fit of the FLMP. The RMSD(b) values averaged .0723, and these values were not significantly different from the observed RMSD values, $F(1, 25) = 3.585$, $p = .067$. This result shows that the FLMP describes the results as well as can be expected for an accurate model.

The analysis of benchmark RMSDs reinforces our dismissal of categorical and holistic interpretations in favor of the FLMP. Given that the FLMP gives the best possible description of the results, it can be claimed that other models could do as well but not any better. Given that the categorical and holistic models (as currently described in the literature) would necessarily be less parsimonious than the FLMP, we conclude in its favor.

The RTs of the identification judgments can be used to test the FLMP's prediction that RT should increase to the extent the facial information is ambiguous (i.e., does not provide stronger support for one response alternative or the other). Ambiguity is defined as the extent to which P(happy) approaches .5:

$$A = 1 - 2(|.5 - P(\text{happy})|), \tag{8}$$

where $|x|$ is the absolute value of x. In this case, ambiguity varies between 0 when P(happy) is 0 or 1 and 1 when P(happy) is .5. An RT averaged across all participants was computed for each of the 35 stimulus conditions and correlated with the A values computed from the average results of the two-choice task of Experiment 1. Figure 8 shows the strong relationship between this measure of ambiguity and RT. There was a strong positive correlation between A and RT ($r = .8289$, $p < .001$). Thus, this RT analysis supports the conclusions reached from the model tests on the identification judgments in that both dependent measures provide support for the FLMP account of the processing of facial affect.

The RTs also provide evidence against the claim that affect perception is categorical and that gradations of affect are not easily perceived within an affect category. If perception of affect is truly categorical, then the time to make a particular categorization should not depend on ambiguity of the stimulus features.

Tests of the models also were carried out using the rating judgments from Experiment 2. The average affect ratings from each participant were fit by both the FLMP and the AMP using the STEPIT subroutine. As can be seen in Figure 5, Equations 3 and 4 of the FLMP gave a good description of the rating judgments. Figure 9 gives the same observed results along with the predictions of the AMP. The average RMSD for the FLMP fit to rating data was .0471, compared with the larger .0756 RMSD for the average fit of

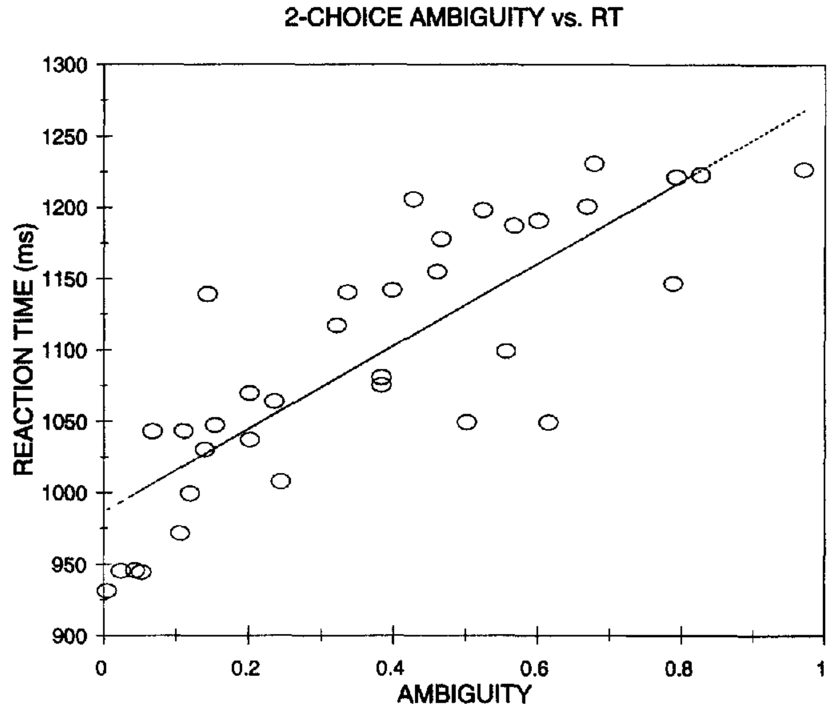

Figure 8. Reaction time (RT) as a function of ambiguity, as given by Equation 8, for the 35 conditions in the two-choice task in Experiment 2.

the AMP. Comparison of RMSD fits showed that the FLMP provided a significantly better overall fit to the data than did the AMP, $F(1, 21) = 136.52, p < .001$.

General Discussion

Our experiments proved successful in addressing the issue of how two features of facial affect are evaluated and integrated to achieve perceptual recognition. Both BD and MD were effective in changing the judgment from happy to angry. In addition, the influence of one of these features was larger to the extent the other feature was ambiguous. These results were well described by the FLMP relative to the poorer description given by the AMP.

The results also are relevant to the issue of whether recognition of facial expressions of affect is or is not in some sense a holistic process. If the FLMP, a model assuming continuous independent features, provides a sufficiently good fit to experimental results, a holistic model would not only have to generate similar fits, but it also would have to be as parsimonious to remain viable. The good fit of the FLMP and the poor fit of the AMP also weaken theories of categorical perception of affect (Etcoff & Magee, 1992), as well as additive models (Huber & Lenz, 1993).

RTs increased to the degree that the stimulus face displayed an ambiguous expression. In accord with the predictions of the FLMP, participants took longer to respond when features were displayed in conflicting directions, when features were only slightly deflected, or when features were missing. Participants responded fastest when features were congruent. Categorical perception cannot easily explain this relationship between RT and feature ambiguity (Massaro, 1987b, pp. 110–114).

Our results seem to indicate that the perceptual processing of facial affect proceeds in accordance with well-established principles of pattern recognition. It is unnecessary to postulate any "special" attributes of facial affect to explain participants' performance. As a case in point, Brown and Dooling (1993) found their effects only in parakeets raised in group environments; an isolation-reared parakeet that had never seen a conspecific (or its own reflected image) did not show any differences when discriminating scrambled or normal parakeet faces. Although data from a single bird can be suggestive at best, this finding seems to indicate to us that face recognition in birds cannot be "holistic"; instead, it requires a component of featural learning through prior exposure in context, just as do other types of pattern recognition and identification. Arguably, a normally arranged parakeet face is just as ambiguous as a scrambled one when viewed for the first time and without behavioral context.

Our experiments show that facial affect perception can be modeled to a high degree of accuracy by a continuous feature model (the FLMP). Additive models (a single-channel model, a weighted averaging model, or a categorical model) were not able to fit the identification and rating judgments as well as the FLMP. Furthermore, the benchmark measure of model fitness revealed that the FLMP provided the best fit possible. Thus, holistic and categorical viewpoints are challenged to provide not only a similarly good description but also one as parsimonious.

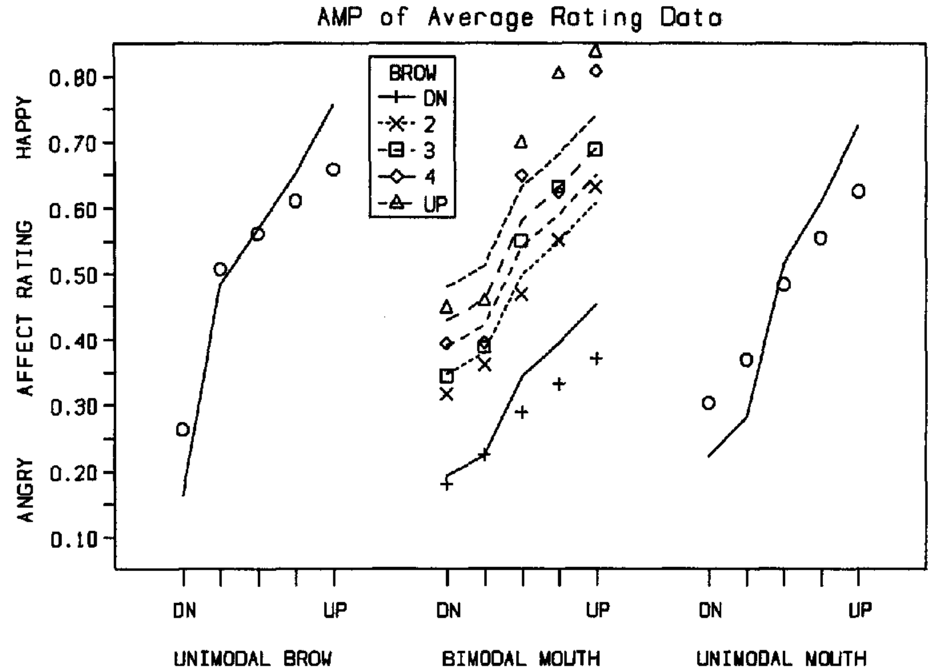

Figure 9. Predicted (lines) and observed (points; *R*[happy]) rating judgments in Experiment 2 as a function of the brow and mouth conditions. Brow DN corresponds to the eyebrows fully depressed and brow UP to the eyebrows somewhat elevated and arched. Mouth DN corresponds to the mouth corners fully curled down and mouth UP to the mouth corners fully curled up. Predictions are for the additive model of perception (AMP).

RTs of the categorization judgments also supported the FLMP in that RTs decreased to the extent the face was unambiguous. These findings show that complex faces are evaluated on a featural basis, and an identification is emitted when total support arriving from multiple sources achieves some critical value; this value is reached more quickly when multiple features are presented that are congruent in their support of a particular affect. On the other hand, if featural information is contradictory, ambiguous, or missing, more time is required before a sufficient degree of support accumulates and a response is emitted (see Massaro & Cohen, 1994).

Given the success of the FLMP across a wide range of empirical domains, its success might not be too surprising. The FLMP has provided an adequate account of the evaluation and integration of sources of information in reading letters and words, in sentence processing, in the visual perception of depth, in memory retrieval, and in cognitive decision making (Massaro, 1987b, in press). It also is noteworthy that most of its success has emerged in the description of speech perception. There has been a long tradition of belief that speech perception is somehow specialized and not amenable to a description grounded in prototypical pattern recognition processes. This belief parallels the belief of many that the perception of facial affect also is specialized. Many of our studies in the speech domain have weakened the foundation of the speech-is-special viewpoint (Massaro, 1987a, 1987b). Although we might predict the same for facial affect, we hope that the success of the FLMP in the current research will encourage other investigators to exploit the use of expanded factorial designs and the test of quantitative models in their research. We believe that this design and analysis provides a microscope that can reveal fundamental processes in the perception of facial affect.

Our research has shown that a standardized set of stimuli that can be controlled through variation of individual features is useful in the exploration of facial affect perception. Previous researchers Ekman et al., 1972 (Ekman & Friesen, 1975; Etcoff & Magee, 1992; Tanaka & Farah, 1993) have not used stimuli that are comparable across studies. The continuing development of the Face39 facial synthesis program in our laboratory promises to provide researchers with a replicable and quantifiable stimulus set for research into the perception of facial affect. With the provision of a coherent type of stimulus and a logically plausible method for varying features within those stimuli, a program of research can be made feasible to explore gender and cultural differences, hemispheric specialization effects, and other influences on the perception of facial affect. Rigorous characterization of features of affect, as well as studies of covariation and facial motion, are needed to implement this research goal. Further studies of affect recognition using expanded factorial designs may shed further light on general processes of pattern recognition. The current work is being extended in our laboratory by performing a same–different discrimination task, which will determine whether participants can discriminate facial expressions within the same affective class less well than they can when the expressions appear to "straddle" an affective class boundary. Further work to explore other features and other categories of affect also should be performed using the paradigm we have described.

References

Allen, V. L., & Atkinson, M. L. (1981). Identification of spontaneous and deliberate behavior. *Journal of Nonverbal Behavior, 5,* 224–237.

Brown, S. D., & Dooling, R. J. (1993). Perception of faces by budgerigars (*Melopsittacus undulatus*): II. Synthetic models. *Journal of Comparative Psychology, 107,* 48–60.

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253–274.

Chandler, J. P. (1969). Subroutine STEPIT finds local minima of a smooth function of several parameters. *Behavioral Science, 14,* 81–82.

Cohen, M. M., & Massaro, D. W. (1992). On the similarity of categorization models. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 395–447). Hillsdale, NJ: Erlbaum.

Cohen, M. M., & Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In N. M. Thalmann & D. Thalmann (Eds.), *Models and techniques in computer animation* (pp. 139–156). New York: Springer-Verlag.

Cohen, M. M., & Massaro, D. W. (1994). Development and experimentation with synthetic visible speech. *Behavior Research Methods, Instruments, and Computers, 26,* 260–265.

Collier, G. (1985). *Emotional expression.* Hillsdale, NJ: Erlbaum.

Cutting, J. E., Bruno, N., Brady, N. P., & Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General, 121,* 364–381.

Duchenne de Boulogne, G.-B. (1990). *The mechanism of human facial expression.* Cambridge, England: Cambridge University Press. (Original work published 1862)

Ekman, P. (1973). (Ed.). *Darwin and facial expression: A century of research in review.* San Diego, CA: Academic Press.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48,* 384–392.

Ekman, P., & Friesen, W. (1975). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings.* Elmsford, NY: Pergamon Press.

Ekman, P., Hager, J. C., & Friesen, W. (1981). The symmetry of emotional and deliberate facial action. *Psychophysiology, 18,* 101–106.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227–240.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21,* 628–634.

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

Harrison, D. W., Gorelczenko, P. M., & Cook, J. (1990). Sex differences in the functional asymmetry and facial affect perception. *International Journal of Neuroscience, 52,* 11–16.

Huber, L., & Lenz, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology, 46,* 1–18.

Kirouac, G., & Dore, F. Y. (1983). Accuracy and latency of judgment of facial expressions of emotions. *Perceptual and Motor Skills, 57,* 683–686.

Kirouac, G., & Dore, F. Y. (1985). Accuracy of the judgment of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behavior, 9,* 3–7.

Massaro, D. W. (1987a). Categorical partition: A fuzzy logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 254–283). New York: Cambridge University Press.

Massaro, D. W. (1987b). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1988). Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General, 117,* 417–421.

Massaro, D. W. (1989). *Experimental psychology: An information processing approach.* San Diego, CA: Harcourt Brace Jovanovich.

Massaro, D. W. (in press). Developing and evaluating computational models in cognitive science: How we read letters and words. In D. N. Osherson & H. Lasnik (Eds.), *Invitation to cognitive science* (2nd ed.). Cambridge, MA: MIT Press.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science, 1,* 55–63.

Massaro, D. W., & Cohen, M. M. (1993). The paradigm and the fuzzy logical model of perception are alive and well. *Journal of Experimental Psychology: General, 122,* 115–124.

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1107–1128.

Massaro, D. W., & Ferguson, E. L. (1993). Cognitive style and perception: The relationship between category width and speech perception, categorization, and discrimination. *American Journal of Psychology, 106,* 25–49.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review, 97,* 225–252.

Sergent, J., Ohta, S., MacDonald, B., & Zuck, E. (1994). Segregated processing of facial identity and emotion in the human brain: A PET study. *Visual Cognition, 1,* 349–369.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46A,* 225–245.

Received November 28, 1994
Revision received September 25, 1995
Accepted October 23, 1995 ■