UNIVERSITY MICROFILMS, INC., ANN ARBOR, MICH.

L. C. Card no. Mic 59- 2173

ROYAL, Donald Cochran. A MULTIDIMENSIONAL ANALYSIS OF PERCEPTION OF EMOTION FROM SCHEMATIC FACIAL EXPRESSIONS.

University of Michigan, Ph. D., 1959
Psychology, experimental

Negative microfilm received from university.

A MULTIDIMENSIONAL ANALYSIS OF PERCEPTION OF EMOTION FROM SCHEMATIC FACIAL EXPRESSIONS

by

Donald C. Royal

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
1958

Doctoral Committee:

Assistant Professor William L. Hays, Chairman
Professor Clyde H. Coombs
Professor Paul Henle
Professor Arthur W. Melton
Associate Professor Anatol Rapoport

ACKNOWLEDGEMENTS

My greatest thanks must go to Dr. William L. Hays for the inspiration and constant encouragement he provided at all stages of this study. His gracious aid and patient understanding proved invaluable in its successful completion. To him and Dr. Clyde H. Coombs, I owe a large debt of gratitude for initiating and guiding the development of the concepts of measurement which form the core of this investigation. It has been a real privilege to be enabled to take even a small part in mapping out some of the new vistas in psychological measurement which Dr. Coombs has opened up for his students.

I am very grateful to Dr. Paul Henle, Dr. Daniel Katz, Dr. Arthur W. Melton, and Dr. Anatol Rapoport for their suggestions and gracious cooperation in this endeavor. My thanks also go to the many students whose cooperation as experimental subjects made this study possible; and to my wife, Pauline, who has rendered immeasurable assistance in every phase of the design and conduct of this study from its inception to its completion.

TABLE OF CONTENTS

Chapter		Page
I.	BACKGROUND OF THE PROBLEM: STIMULUS CHARACTERISTICS	1
II.	BACKGROUND OF THE PROBLEM: RESPONSE ASPECTS	23
III.	PRELIMINARY EXPERIMENTS	41
	The First Experiment	41
	The Second Experiment	44
IV.	THE SPECIFIC PROBLEM	49
	Introduction	49
	Hypotheses	51
V.	PROCEDURES AND RESULTS	54
	Study I	54
	Study II	63
	Study III	76
	Reliability	79
VI.	SUMMARY AND IMPLICATIONS	86
	Implications for Future Study	87
APPENDIX A		95
APPENDIX B		96
APPENDIX C		97
APPENDIX D		98
APPENDIX E		98
APPENDIX F		99
BIBLIOGRAPHY		100

LIST OF TABLES

Table		Page
1.	Distribution of Judgments for a Rough Scale of Poses	28
2.	General Dimensions of Judgment of Emotion from Schematic Facial Expressions	59
3.	The Degree of Consensus Among Judges on Each Triad Order	62
4.	Dimensions Obtained from Groups Instructed to Judge on Specific Attributes	68
5.	Similarity of Dimensions: Specific Attribute Opposites	71
6.	Similarity of Dimensions: General with Specific Attributes	72
7.	Similarity of Dimensions: General with Specific Attributes	73
8.	Similarity of Dimensions: General with Specific Attributes	74
9.	Primary Attributes Ordered as to Similarity to Each General Dimension	75
10.	Scales Obtained by Matching the Dimensions of Human (Male) Emotion Found in Preliminary Study, and Degree of Correspondence	78
11.	Individual Dimensions	79
12.	Correspondence of General Group Dimensions with Individual Dimensions	81
13.	Average Correspondence of General Group Dimensions with Individual Dimensions	82
14.	Average Correspondence of General Group Dimensions with Individual Dimensions Non-maximal in Similarity	83
15.	Average Interagreement between Individuals' Dimensions	84

Chapter I

BACKGROUND OF THE PROBLEM: STIMULUS CHARACTERISTICS

Facial expression was probably studied originally for its value as an indicator of underlying emotion. Yet, in most of the studies on facial expression, there is surprisingly little consensus as to what emotion a given expression signifies. Each facial expression stimulus serves as a cue, having both common and unique components of meaning for the members of any judging group. The presence of this unique meaning component for a given judge tends to reduce apparent consensus. A large part of this unique or individual variance should be related to the "social context" to which a human face is referred; such social context is not a part of the stimulus itself, but is tacitly supplied by the judge to supplement the objective stimulus.

If one is interested in the common, consensual, factors underlying judgment of emotion from facial expression, it is, of course, necessary to minimize this variance attributable to social context supplied by the individual judge. This additional, and possibly inconsistent, social context may be reduced by the use of schematic faces as stimuli. The dimensions of judgment inferred from the use of schematic faces showing controlled variation of cues should provide a means to better understanding of the consensual, culture value of such cues in inferring emotion.

Studies of facial expressions of emotion might be conceptualized as studies of a communication process. That is, the observer characterizes a facial expression, intended to depict a particular emotion, by assigning a verbal label to it. It has been frequently assumed that this form of communication has precision and reliability even though there is a large amount of evidence to the contrary. To be sure, the emphasis has usually been on evaluating the observer rather than the classification system. Nevertheless, the studies to be examined below show that agreement is notably poor among observers over what word labels should be used to describe each emotional expression.

Typically, the early studies initiated by Piderit (38) and Darwin (14) in the 19th century employed the presentation of several stimuli such as photos, diagrams, or poses of facial expressions to judges whose task it was to label the emotion being expressed. The judges achieved varying and often disappointing degrees of success - as measured by the accuracy with which the emotions were judged, i.e., the degree to which the judgments made by an individual corresponded to the group norm or to the intentions of the poser or experimenter.

One series of studies, particularly relevant to the present research, employed artificial faces with interchangeable features. Boring and Titchener (4) used a face model with interchangeable components and found that all expressions were accepted as indicative of some emotion even when the features were inconsistent or contradictory. In another study, Buzby (9) allowed the use of any of eighteen labels to categorize each of five Boring and Titchener model faces. The average accuracy figures (percentage of judgments agreeing closely

with the intended emotion) for his judges, on each face, ranged from 5 to 57 per cent - not as high a degree of consensus as one might expect from simple unambiguous stimuli. Fernberger (19) found that suggesting a "false" label (the name of an emotion other than the emotion intended to be expressed) to a judge, on the average, increased the judge's tendency to use that label to characterize that expression from less than 10 per cent to more than 70 per cent. Perhaps drawings or diagrams are somewhat inadequate to convey emotions, being too divorced from the total stimulus situation. Yet judgments from photos of posed expressions are little better. Langfeld (34) used Rudolph's poses (40), and had judges give name descriptions of each. The most successful judge gave 58 per cent correct names (those names agreeing closely with the actor's intentions). The least successful judge gave 17 per cent correct. This same set of judges accepted from 16 per cent to 47 per cent "false" responses suggested to them.

Obviously the labels available to a judge exert a considerable effect on his categorizing behavior. The mere existence of a category or label induces an expectation of an appropriate emotion. Thus the emotion acquires a reified existence and properties all out of proportion to actually observed properties. It may be possible to secure better agreement by returning to simple observation.

There are actually several difficulties involved in the process of inferring emotion from facial expression. Even though individual experimenters have often taken into account one particular experimental refinement, few have systematically avoided the remaining difficulties.

There is no precise connection between facial expression and emotion any more than there is between words and emotion or expression. Insofar

as some attributes other than emotional ones can be inferred from facial expression, and as long as there are components of emotion varying independently of facial expression, one cannot expect a precise fit between emotion and its facial expression. Therefore, in any systematic study, one must try to eliminate as many as possible of the non-relevant variables, such as, for example, observer-supplied context.

The judgment of emotion from facial expression is a special case of social perception and is presumably subject to all the errors of such perception. Studies of social perception (5) reveal that certain characteristics of a judge may affect his perception of social objects. Degree of self-insight, empathy, "openness," aesthetic sensitivity, detachment, intuitiveness, as well as social adjustment and ethnic group membership all seem to be relevant factors. However the lack of comparability between these concepts renders quite hazardous any particular conclusion - especially since the evidence is somewhat inconsistent in intuitive interpretation. For example, degree of "openness" and "introvert" type of social adjustment both favor accurate perception (5), but may not intuitively seem related, at least within a single person. Concerning experience as a correlate of ability to judge others accurately, Bruner and Tagiuri (5) assert that little systematic testing is available to prove or disprove the point. The main difficulty seems to be a lack of explicitness in stating what is involved in "experience."

As for general intelligence, it seems that lack of it may adversely affect accuracy of judgment but high intelligence is no guarantee of good performance (5). Of particular import is the notion, advanced by G. W. Allport (2), that a subject may have difficulty in judging one

more "complex" than he (the subject) is. Though Bruner and Tagiuri (5) in their review of this literature found little evidence for age differences in general social perception, age differences could be important at lower age levels where social concepts are still in a fluid stage of formation. In any case, it is probably not wise to assume a high level of complexity or intellect in a judge; it seems more feasible to simplify the stimulus situation to its essential aspects.

What characteristics of the judge aid particularly in the recognition of emotion? There is evidence of very early ability to distinguish certain emotions in others (8, 48). Many of the early reactions by the infant to contorted faces seem, on further analysis, to be of a rather non-discriminatory nature. However, after several months, he may begin to react more specifically to certain stimuli. Gates (22) tried to ascertain at what ages recognition of different emotions becomes possible. Using a criterion of 50 per cent correct responses, he found that children of three years of age could recognize the Ruckmick poses (39) for laughter. Recognition of pain, anger, fear, horror, surprise and contempt followed in steps of about two years. Recognition of contempt reaches criterion at 14 years of age.

Since it seems likely that the ability to judge emotion in others develops with experience, F. H. Allport (1) attempted to train subjects to improve their recognition of emotions. He showed judges 14 Rudolph poses, and obtained an average score of 49 per cent correct. He then lectured intensively on facial anatomy and the specific contributions of the muscles in emotions, repeated the series of faces and obtained

an average score of 55 per cent correct. Although initially low scores improved much, the initially high scorers improved very little.

Quoting Bruner and Tagiuri,

> As to sex differences, the literature is somewhat contradictory. F. H. Allport (1) and Guilford (23) have reported no differences in judging the Rudolph pictures. Jenness (27), using a larger sample, obtained superior scores for women. Kanner (28), using the Feleky series, found that men excelled slightly. Fernberger (19) adds to the neutralist position a study indicating no sex differences on the Boring-Titchener model /with interchangeable components/.... Coleman (11) reported no sex differences using movies of laboratory elicited emotions. (p. 639, 5)

Age level then appears to be the only characteristic of a judge that has been demonstrated to affect accuracy of recognition of emotion. Since poor results are still obtained when mature subjects are used, one must look elsewhere for the sources of inaccuracy.

The discussion of the relation of other components of emotion to facial expression will be left until later; studies of the perception of emotion itself will be examined first. Within an experimental setup, the stimuli (photos, diagrams or real people) employed have varied widely with respect to the amount of potential information they contain. For example, the person (or photo of a person) may be acting spontaneously in a real-life situation (36), or spontaneously within a laboratory situation (33) or merely posing (21). It might be expected that the more artificial the situation, the less accurate the recognition of emotion. However, Landis (33) tried many different laboratory situations to produce "spontaneous" emotions. A later analysis of the photographic records revealed very poor discrimination.

Munn (36) used candid shots in "real life situations" from Life and Look magazines and found marked improvement in judgment as more

cues were added. Of course it does not follow that, just because actors' poses may be highly conventional, that the faces of everyday folk are spontaneously expressive. In judging pictures of "real life situations," it does not suffice to recognize an emotion merely by judging it as appropriate to the situation. Under conditions of social contact, it is possible that expression may be more of a social response than a spontaneous expression of feeling. The difficulty is that here there is no independent identification of the emotion or situation being judged. To a great extent this difficulty plagues all studies which purport to have the judge recognize a particular emotion. The study which probably comes closest to using spontaneous expressions is that of Sherman (47). The difficulty is to secure known conditions of emotion. Sherman applied the Watson stimuli for fear and anger to a series of infants and took motion-pictures of their responses to these stimuli, as well as to pain and hunger. When only the response pattern was shown, judges had great difficulty and poor agreement as to emotion. When the complete film, including stimulus situation, was shown, much closer agreement was found. Frijda (20), however, found that the addition of cues by showing a temporal sequence (a movie of acting the emotion out) contributed enough negative cues to almost balance out any advantage of the addition of cues. Of course, when such situational cues are added, facial expressions per se are no longer the variable being investigated.

As Fernberger (19) puts it, "If a stimulus situation is indicated, the emotional state is judged in accordance with the facial expression."

Things other than emotion, then, can be and often are, inferred from a face. These other inferences may be extraneous to our present concern

but often do exert considerable influence as contextual constraints, much as situational data drastically influence the perception of emotion. Insofar as these inferences are present in all facial expressions, their presence must be endured. However, as long as they are systematic (whether by direct control or by balancing them in such a way that their relevant contribution is equivalent throughout), they are not the source of the lack of consensus so prevalent in perception. It is the idiosynoratic features that must be minimized.

G. W. Allport (2) lists several variables which affect a person's judgment of others - similarity, complexity, detachment, esthetic attitude, etc. It is clear that the presence of strong idiosyncracy in particular stimuli may produce widely varying effects, quite irrelevant to the emotion depicted, on different perceivers.

As an example, consider just one of these factors - the degree of similarity between the judge and the person being judged. Hypothetically, it might be expected that all other things equal, a person is more accurate in perceiving someone who strongly resembles himself. Although any of a number of quite independent factors may be responsible for this result, this hypothesis has frequently been substantiated (5).

The problem attains great complexity when photos depicting different people are used in the same set of judgments. Evidence on the formation of general impressions in interpersonal perception indicates we are more likely to be judging the persons themselves than the emotions being expressed. Thus, two components of judged emotion may be identified - the person as he characteristically <u>is</u>, and the <u>transitory state</u> of his emotions.

It may prove rather difficult to separate these two components, as the perceived characteristics of an individual will always affect judgments of the emotions he is capable of expressing. Yet considerable simplification can be achieved if such "trait" characteristics are kept constant, and only the relevant expressive characteristics are varied.

The Ruckmick (39), Feleky (17) and Frois-Wittman (21) series of facial expressions can be considered as attempts to maintain comparability within a series of judgments by minimizing variation in non-relevant features. Their systematic employment in studies of facial expression constitutes perhaps the most significant advance in experimental procedure prior to 1935.

It has made possible a certain degree of comparison, necessary to achieve generality. Small variations of undetermined effect are still present, though. Such variations in lighting effects, shadows, angle of view, size of face and its distance from camera are often considered important in dramatic portrayals of emotion by professional cameramen. There has been little systematic investigation of these factors although Secord (45, 46) has found that small variations in brightness or texture of skin perceived from photos may exert considerable effects on impressions of traits, including temperament or affect-related variables. Such effects as these are only compounded in their effect when the perceiver must keep shifting his judgmental basis with the sequence of differential stimuli.

Secord (46), in an extensive series of studies, has investigated what impressions of personality can be inferred from photos of faces. He differentiates between structural and expressive cues, and seeks to

study the former intensively while minimizing the effect of the latter. To accomplish this he eliminates pictures which seem to convey great degrees of emotion. Of course, this does not eliminate the effect of expressive cues since they are still present and are often quite inseparable in operation from more permanent structural characteristics. But it is presumed that their operation does not predominate over structural cues in producing impressions of personality. Although the dichotomy is assumed more for convenience than for descriptive precision, one may assume a corresponding difference in the type of concept conveyed by these cues. The types of things inferred from structural cues are presumably of a rather permanent nature - personality traits, character traits, whereas inferences from expressive cues refer to more transient qualities - moods, feelings, etc. There is, however, considerable overlap between these distinctions. Many of Secord's consensual personality impressions refer to characteristics of an individual which may be regarded either as permanent or transient - depending on one's own preference, past experience and criteria of temporal reference. It is questionable whether such a rigid dichotomy can ever be enforced, but more steps directed towards standardization of stimuli are essential before this goal can even be approached. The stimuli should be comparable in all respects save the type of cue under investigation, which may be varied systematically. This type of design should enable one to make more precise assessments and inferences about the particular area with which we are concerned.

Investigations of physiognomic characteristics of an individual have in the past concentrated perhaps more on permanent characteristics

than on the cues of transitory emotional states. In this respect Secord differentiates between two forms which investigations of physiognomy have traditionally taken: 1) Appraisal of physiognomy by taking various physical measurements of the facial features in the attempt to relate these measurements to known personality characteristics of the individuals measures; or more frequently 2) judgments of personality traits from physiognomy and their comparison with known personality traits.

Regarding the first, such a procedure has at least two noteworthy disadvantages: a) variations in the physical size of the faces (as well as the angles subtended) from photograph to photograph would necessitate the use of some complex proportional method of measurement; b) many important physiognomic characteristics, for example, eye depth, are not easily amenable to linear measurement. Ignoring the rather obvious difficulties in the measurement of the other component, personality traits, one has at least some reliable instruments for their detection, <u>independently</u> of physiognomic judgments. However, in judgments of emotion there is an even greater disadvantage since no reliable instruments of interpretable validity exist. In a sense, the physiognomic judgments themselves must often serve to validate other measures. This situation is primarily due to the very transitory nature of emotion one is attempting to study, there being few other indicators which reflect emotional states as quickly as expressive behavior. Since the dispositional nature of an emotion is of short and uncertain duration, satisfactory situational measures are extremely difficult to devise. This problem of validity has been and will remain a thorny one.

Secord (46) avoided this problem of validity by comparing perceived physiognomy with impressions of personality without regard to any characteristics of the stimulus of the individuals besides those actually presented in the experimental situation. He secured considerable consensus within several clusters of personality impressions (that is, personality trait-ratings could be analyzed into several clusters within which trait-ratings correlated highly). These six clusters were labelled by Secord as a) warmth-tolerance, b) moral-social respectability, c) arrogance-intolerance, d) forcefulness, e) aloofness, and f) meekness. Each cluster then included, also, some physiognomic traits most common to that cluster. For example, cluster E (aloofness) had for the most characteristic personality traits - poker-face, hostile, boorish, sly expression, quick-tempered, and conceited look; and for the least characteristic personality traits - likeable, friendly, honest face, cheerful appearance, sense of humor, trustful; and had for common physiognomic traits - dark complexion, oily coarse skin, dull eyes, facial tension, heavy eyebrows, average age face, medium eye depth, straight mouth.

This type of analysis is interesting because it affords one avenue of escape from the problem of validity, since in neither case are the ratings claimed to extend beyond impressional status. Such a procedure, however, stops short of the goal of establishing an observational basis for judgments of emotion. One danger in dealing with these two perceptual dimensions is that impressions purporting to be of physiognomy may become as evanescent as impressions of personality. That is, so many physical cues may contribute to the formation of the impression that we perform an injustice by selecting only one as responsible.

Secord's distinction between physiognomic traits and personality traits in practice seems rather arbitrary, as for example - the selection of "sly expression" as a physiognomic trait, and "facial tension" as a personality trait is not entirely obvious. This difficulty of distinction is further accentuated when the traits in question have no independent accessibility to observation. The procedure then is dependent on the experimenter's selection of trait names and what the subject understands is meant by them.

The selection of trait names has in the past been far from arbitrary - often being dictated by quite extraneous non-empirical considerations. Allport and Odbert (3) studied approximately 18,000 terms in the English language designating distinctive and personal forms of behavior. They suggest two major influences on the formation of such terms. First, a desire to represent by name such mental processes or dispositions of others as can be determined by observation or inference. Second, "the tendency of each social epoch to characterize human qualities in the light of standards and interests peculiar to the times. Historically, the introduction of trait-names can be seen to follow this principle of cultural (not psychological) determination to a striking degree." Of course, such predominant cultural influence is to be expected, but can be quite deleterious to accuracy of usage when the particular cultural values or referents change. In any case, externally imposed influences can greatly distort the psychological phenomena.

Concepts originating out of cultural ideals or social judgment may and often do become guiding principles adopted by individuals. Character judgments probably become accepted as "traits" in this manner. Allport

and Odbert caution about the validity of these terms, especially those based on evaluation:

> The theory we present then holds trait-names are symbols socially devised (from a mixture of ethical, cultural and psychological interests) for the naming and evaluation of human qualities. Some of these terms are obviously censorial and as such have little utility for the psychologist. The non-censorial terms, however, are significant, for their common usage establishes a presupposition that some human beings possess actual dispositions or traits roughly corresponding to these symbols. There are, however, many more traits than any list of single names would indicate, for we often find neologisms, phrases, and metaphors called upon where trait-names are insufficient.
>
> In scientific work no single trait-name can be accepted with assurance....until its correspondence with a true trait has been experimentally or clinically established. Traits cannot be called forth by fiat; they must be discovered. (p. 20, 3)

In spite of their apparently great linguistic and cultural determination, the accurate communication of these concepts to the judge is even more hazardous. The terms are often overly specific in their reference, and the remoteness of these referents and their frequently neologistic usage by the experimenters does not aid in their consensual comprehension by the perceivers. Even if agreement on an area of meaning could be reached, the impressions actually formed, the cues which mediate them,and their influence on perception,may differ widely. Unfortunately, the definiteness of form conveyed by the existence and use of such terms is a poor indicator of their actual unreliability.

What is found is a kind of commonality of verbally-induced impressions within each of several subjects, with no assurance that the same physical characteristic is mediating these. But this attainment of a measurably high agreement is in itself a significant achievement, especially when the abortive attempts to achieve agreement on emotion are

reconsidered. However, there may be basis for expecting more agreement with respect to personality, as traits of personality are more permanently conventionalized by social habit. There seems to be a strong tendency within society to categorize individuals permanently at certain levels.

Returning to more transitory states, it cannot be denied that there may exist other components of emotion besides facial expression. There is abundant evidence indicating the presence of feeling or changes of feeling without a concomitant variation in facial features. Immobile features, especially in certain cases of schizophrenia need not always indicate an absence of feeling or a "fixed" emotion. A person can often learn other ways of expressing his feelings, or suppressing them if necessary. The literature of psychoanalysis is rife with examples of this. Not all examples need be pathological for much of the entertainment industry depends on an actor's ability to express feelings in a highly visible form, even if he does not experience the feelings himself. Landis' results (33) show the often considerable influence a social situation may exert on expression, if not on emotion itself.

Certain persons, especially actors, can probably "fake" the expression of emotions, that is, go through most all the visible motor aspects of expressing an emotion, without "really" experiencing the emotion, in spite of the widespread influence of the James-Lange theory. Much fruitless controversy in the area of emotion has centered on this question, of whether expression is a necessary or sufficient component of emotion, which, in its usual forms, is not susceptible to empirical test.

No claim will be put forth here as to the essential core of emotion, if there is one. No references to unobservable and unique individual

experience should form the basis of our analysis of emotion. All these early attempts presuppose another identification of emotion, independent of its expression. This identification is rarely possible for an external observer to attain, especially in everyday life situations (52).

Our culture doubtlessly has some kind of tendency to notice recurrent events and establish conceptual schemes for describing, explaining or channeling them. It is only when we use these schemes to declare that there can be nothing else of significance but what we have subsumed under our particular scheme that one ponders their misuse. When we examine the data of facial expression of emotion there seems to be considerable evidence of social effects, perhaps even to the extent of producing social stereotypes of expression.

We often speak of certain kinds of expressions as typical of certain cultures, or as having different meanings. Ruesch and Klees (41) cite the example that "hissing in Japan is a sign of polite deference to social superiors, whereas in England as well as in America it is the exact opposite - a sign of extreme disapproval." Ruesch and Klees go on to discuss gestural expression as a form of nonverbal communication:

> The organization of the communication network and the modes of communication are fitted to suit typical and recurring situations. The typical element is usually not so much a single gesture or a single action, but the patterning of various symbolic and adaptive expressions. The pattern can often be identified on an abstract level only, either in terms of timing, rhythm, repetition, or in terms of spatial arrangements or contest. For example, variations in climate, the density of population, the prevalence of either an agricultural or an industrial economy, the availability of raw materials, the historical past, and many other factors, will all determine in part the varieties of communication. In any culture each person is prepared, therefore,

> through education to assume a place in the communication network of his particular group, and the cumulative body of experience of generations reflects itself in the expressions of the individual. This particular thesis - the subservience of movement and gesture to long-range values - is illustrated below with examples drawn from five different cultures. Gesture among the Americans is largely oriented toward activity; among the Italians it serves the purposes of illustration and display; among the Jews it is a device of emphasis; among the Germans it specifies both attitude and commitment; and among the French it is an expression of style and containment. (p. 22, 41)

Of course, many of the observations used to support this point are based more on personal impressions (which may not even be self-consistent) than on experimental data. But it is wise to keep in mind that facial expression can serve other ends besides those of emotion. There may be culture influences but it is wise to question, via empirical investigation, whether the culture stereotype is more in each culture or more in the mind of the observer when he speculates about the culture.

Certainly some specialized modes of communication may develop as a result of some specific cultural influence as in the development of a one tone whistled form of the Spanish language on an extremely mountaineous isle in the Canary Islands (10). These need not reflect any new meanings, although specialized meanings may accrue as an outcome of having special forms of expression. There may be retention of both specialized and more universal meanings, distinguishable by context. Or there may simply be some or all gestures overemphasized without any especial extra significance.

In other words the few empirical studies attempted have generally focused on <u>differences</u> in expression between cultures, without determining

the degree of similarity, either between or within the cultures.

In an extensive study of emotional expression among the Chinese, Klineberg (30) points out a considerable number of differences from Western expression. For example, in a Chinese treatise on acting, "To draw one leg up and stand on one foot means surprise." In Chinese novels, some emotions-especially fear and sorrow-seemed to be expressed similar to Western forms. "They stretched out their tongues" is frequently used as an expression of surprise. A smile means contempt or anger much more frequently than it does in Western expression. Although this study was largely based on selected literary materials, it does suggest possible cultural diversities in expression. However, suggestive as they may be, such studies have shed little light on the extent of culturally universal expressions of emotion.

Although there may be cultural differences in expression of emotion, it need not follow that these differences are perceived differently. That is, a perceiver may be aware of both general and specialized meanings and be able to use either as the situation demands. In this respect, it is wise to recall that there are two aspects to the perception process - the stimulus or the expression which one perceives, and the perceptual response itself. The kinds of emotion expressible and the kinds of emotion perceivable may not be coincident. If the interest is more in the perception of emotion, rather than its expression, once a stimulus representation with reasonable prospects for stabilization within a particular culture is selected, the focus may be put on perceptual response variables.

No definite conclusions can be made at present on the extent of commonality between cultures in the perception of emotion. But not

every cue is pre-ordained to have a single inherent expressive significance for emotion. The significance of a cue lies in its perceptual consequences. One should not be disturbed if these be extensively influenced by social conventionalization, since it is those conventions which influence and determine perceptions which are to be discovered. The influence of the social milieu on perception might be less palatable if we were inferring directly our own emotions, where social influence may not necessarily be a factor. But social influence on categorizing behavior and the influence of categorization on perception is relatively incontestable. In any case, a social or cultural milieu of some kind is logically inescapable in any observation. The difficulty is that in any social situation many unknowns are introduced. Perhaps one of the strongest reasons for using photographs is that they eliminate the reaction of the poser or expressant to the perceiver and perhaps minimize the reaction of the latter to the former.

It is possible to control the stimulus situation to an even greater extent by using schematic faces as stimuli. Then, it is possible to circumscribe and sample the ecological range of stimulus variation, insuring a representative experimental design. However, the ecological distribution of possible stimulus variations cannot be expected a priori to conform to the distribution of physically possible stimuli.

To some extent, studies employing diagrams (4), generally drawn by an artist, or consisting of a combination of interchangeable features, have avoided the problem of stimulus situation and have focused judgments on the facial expression itself. Yet these diagrams are usually in the

form of a highly personalized drawing, including many characteristics not particularly related to expression such as beards, "hair-do," eyelashes, forehead shape, jaw shape and ear shape. These all contribute to the formation of a general impression - an image of a unique and highly stereotyped character. This impression may be the judges' conception of the "causal texture" of the individual, i.e., a judgment of general character may influence the kinds of emotion deemed most appropriate or likely to be consistent with that character. It is necessary to remove as many as possible of the non-relevant variables which contribute to the formation of a particular character impression, especially where this impression may be idiosyncratic (differ from one judge to another).

In order to be sure of doing this, however, one needs to have some knowledge of the range of variables relevant to such perception. Thus, if it is possible to circumscribe and sample the ecological range of stimulus variation, a representative design can be insured. It must be emphasized, however, that the ecological distribution of possible stimulus variations cannot be expected a priori to conform to the distribution of physically possible stimuli.

It might be expected that some sort of estimate of the frequency of occurrence of particular expression, or its observation, could be found. However, in this realm, concepts of frequency seem a little too abstract. Instead of being a primary characteristic, it is only with extreme difficulty (in the preliminary studies discussed in Chapter III) that a perceiver is able to compare relative frequencies or even estimate absolute frequencies of either an emotion or its expression. The whole problem of

the perception of probability and frequency of occurrence is beset with anomalous results, in spite of the rather objective stimuli customarily utilized. When more subjective and complex stimuli are involved, direct estimation of frequency is even more mediated by partial aspects and distorted by idiosyncratic influences.

To determine the maximal range of emotional expressions in the face, it is perhaps most defensible to take the most discriminable (with respect to judged emotional states) facial expressions. These need not necessarily be the faces with the maximal variations in cues. The second preliminary experiment is an empirical investigation to determine the most discriminable expressions.

But first, the most criterial cues active in facial expression of emotion must be selected from among all those cues possible. Brunswik and Reiter (7) designed a series of schematic faces involving systematic variation of the ecological variables: forehead, distance between eyes, and height and position of nose; "tacitly guided by principles of representativeness." The original aim was to investigate the physiognomic impression value of a number of prominent features of the facial geometry including their combination in patterns or "Gestalten." The experimenters chose three possible levels of strength for each of the experimental variables so as to enable the design to bring out a suspected curvilinearity. This type of design was an informal attempt to achieve representativeness and a "normal"-looking standard was used in defining the center strength position. Although Brunswik's choice of stimulus variables perhaps reflects structural rather than mobile characteristics which express emotion, his experimental design merits further consideration, chiefly because of

its success in combining features of both systematic and representative design.

Further investigations (discussed in 6) have been carried out on these same basic schemas in attempting to enrich the cues in the schema and determine the relative contribution of each.

The impression value of the eye region was investigated by Kremenak (31). She varied the size and shape (thickness) of the eyebrows, their distance from the eyes, and the shape of the eyes, the position, size and shape of all other facial features being held constant. The shape of the eyebrows had the greatest expression value, followed by the distance of the brows from the eyes and lastly the shape of the eyes. The attributes of impression on which each face was rated consisted of rather mixed emotion and personality-type impressions, e.g., good, sincere, careless, extraverted, gay, intelligent, energetic, etc., and, therefore, structural cues may be less criterial in the study of emotion.

Winkler (51) used the same dimensions of impression to study the influence of variations in the shape, height and angle of incline of the mouth. Of these the shape of the mouth had the greatest expression value. Further studies have been conducted to study the impression value of hairline, hairstyle, beard, mustache and other characteristics which are either irrelevant as relatively immobile characteristics or are highly idiosyncratic.

The observations presented here seem to indicate the preferability of employing a well controlled set of schematic faces as stimuli. This procedure aids in avoiding the effects of idiosyncracies of social experience which more complex stimuli may evoke in some observers.

Chapter II

BACKGROUND OF THE PROBLEM: RESPONSE ASPECTS

What response measures are selected for study of the perception of emotion from facial expression pose an important problem. Verbal response measures are often misleading as to their precision and standardization. If verbal responses are to be employed, one must analyze and order the observer's "errors" or have the observer himself do this within the experimental situation. The latter alternative usually involves the a priori assumption of a well defined unit of measurement and imposes the dimensions on which one may judge.

Since perception may be defined as a response, the nature of the responses available to a perceiver may affect what he "perceives" (or, more strictly, what an experimenter can infer about the perceptive process). Verbal responses are most commonly used, both in every day use and within psychological experiments. There are many possible forms of verbal response: 1) expressive utterances and cries; 2) descriptive words; 3) simple statements composed largely of 1 and 2; 4) more elaborate propositional statements; 5) allegorical material.

Descriptive words or labels have been most widely used for research on perception. There are probably many reasons for choosing to study labels, among them being a greater chance for uniformity of usage and a greater degree of comparability within the set of possible responses. The meanings of words are often considered less variable than meanings of

more complicated or less standardized constructions. Thus the meanings may be better determined (explanation of the process whereby meaning is acquired is facilitated), and presumably more people are aware of this "fixed" meaning. Two disadvantages of the use of verbal labels are that 1) a single observer does not always use a label in the same way, using a standard meaning. Those words whose meanings are so restricted often drop out entirely from a speaker's vocabulary. Those words retained generally are wider in reference - often achieving the status of a handy word to be used on almost all occasions - as when one might always answer "bushed" or "lousy" when asked how he feels. 2) The advantage for analysis in having "fixity" of meaning may actually be a disadvantage in description itself.

Advocates of new forms in art, music and poetry have long contended the inadequacy of our present verbal concepts with which we try to categorize affective phenomena. In the past scientists may have tended to dismiss such testimony as originating from a professionally entrenched point of view, and therefore to be expected - regardless of any descriptive validity. But the limitations of verbal descriptions are materializing within our own field. Clearly if agreement cannot be reached as to what is seen, our descriptive vocabulary is deficient in some respect. It may be that the terms do not fit the "events" or that so many terms fit the events that it is improbable that perceivers will use common terms. Furthermore, each term may fit so many "events" that it is completely dependent on environmental constraints, or sequential position in usage, to govern its elicitation, regardless of whether it is the most apt term that could be used.

It is notable that early experiments often employed rather precise, often highly intellectual terms which perceivers could use to categorize facial expressions of emotion. The perceivers' inability to use these terms had to be taken then as inability to perceive emotion, as these were often the only measures employed. When perceivers were allowed to use their own terms to categorize, "agreement" improved considerably. To arrive at any high degree of agreement, though, it was often found necessary to count as being in essential agreement different terms which were felt by the experimenter to mean the same. That is, it might be felt that to count a judgment of "happy" as "correct," and a judgment of "joyful" as "incorrect" was going too far and some allowance should be made for "near hits" or the use of closely similar terms. The difficulty here is that many assumptions are being made in this process and it is doubtful at this stage whether they are justified.

Of course this leap must be taken at some stage in analysis. One must code the phenomena into some kind of meaningful forms if one wishes to communicate about it. But one must also select those means of codification that do the least injustice in distorting the original impressions. Hence two problems are evident, concerning 1) the nature of the discrimination demanded, and 2) the nature of the categories utilized, whether by the perceiver or the experimenter. The early studies almost always required the perceiver to classify a facial expression by assigning it a label from either a preselected or a spontaneous set. The meaningful form that was generally used to characterize the experimental results then was to compare the perceived emotions with the "actual" emotions, and arrive at an accuracy measure.

In most studies on facial expression it must be remembered that the criteria of classification are highly arbitrary. There is little reason to assume that the experimenter's judgments or the actor's intentions constitute an ideal standard. Later studies have often taken the majority agreement among judges as a standard. Yet all these studies assume that there is some adequate category to characterize each emotion. To reject as "errors" all alternative classifications seems highly arbitrary to say the least and perhaps is the major reason for the lack of progress in this area. Hulin and Katz (26) avoided this mistake by discarding labels entirely and asking for judgments of which pictures belong together. But a method for combining these into a meaningful form was still lacking.

There are several reasons for this impasse. There is actually little evidence to support the view that the labels have equivalent meanings for all subjects, or that they are equally inclined to use the same categories to classify their judgments. Consequently,in general,it is seen to be hazardous to assume that either introspectively based (validated) verbal labels are adequate descriptions of the emotions (thereby excluding any unverbalizable aspects) or that certain feelings are invariably associated with certain labels. If we assume we are more interested in the basic dimensions which underlie behavior, than in the communication process and language response itself, then there are alternative models that may be assumed.

Bruner and Tagiuri (5) point out that the basic problem here seemed to be the discovery of a method of scoring which would discriminate between errors which were greater or lesser in magnitude. Recognizing the

fallacy of rigid categorization with its concomitant "errors," Woodworth (53) attempted to relate descriptive categories to more basic dimensions by constructing an ordinal scale from data previously collected by Feleky (18). In the original experiment, 100 observers judged 86 poses (photographs of an actress posing for various emotions). The observer was furnished with a list of emotion names but was free to use any word or words that seemed most appropriate. Although a great variety of names are used to characterize any single pose, most responses are roughly synonymous and seldom overlap much with the list of responses to certain other poses. Woodworth was able to use this kind of information - absence of overlap, to construct an ordinal scale of six categories. Each expression was assigned to the category in which it was most frequently classified. A scale of discriminability was constructed by further analysis of "errors of discrimination" (failures to assign an expression to its modal category). He hypothesized that maximal "error" in sorting stimuli should occur between neighboring categories, with "errors" diminishing as categories become further separated. Several categories between which discrimination was minimal were collapsed into single categories. Assigning a numerical value to each category range on the scale, Woodworth was able to re-analyze data obtained with other sets of pictures and obtain reliable results consistent with his scale.

By this approach, Woodworth has perhaps solved the problem of what to do with the data. Of course there remain unsolved methodological questions about which he is not explicit.

> It is a likely hypothesis, therefore, that good correlation exists between the poser's intention and the observer's judgment. This hypothesis will be confirmed

> if we can find an order of the emotions such as to show a good correspondence between pose and judgment - provided also the order is reasonable on other grounds. If we can arrange the emotions in a single series, with similar emotions adjoining, we can obtain a rough measure of the error of each judgment. Probably the whole variety of emotions cannot be arranged in a linear series, but 10 important ones are so arranged in the following table. The order is adjusted to the votes of Feleky's 100 O's, so as to make the frequencies run a regular course in both vertical and horizontal lines, with more overlap between adjacent than non-adjacent columns and rows. In other words the order is adjusted to give the maximum correlation between pose and judgment, and so to test the hypothesis that a good correlation exists. (p. 251, 53)

(The table from Woodworth's study is reproduced here as Table 1.)

Table 1.

Distribution of Judgments for a Rough Scale of Poses
(Data from Feleky, 1924, 18)

Judgments	Poses									
	Love	Happiness	Mirth	Surprise	Fear	Anger	Suffering	Determination	Disgust	Contempt
Love	10	2		1			1			
Happiness	68	63	51	4		1	1			
Mirth	9	27	49					3		
Surprise	1	1		77	5	7	1			
Fear				6	66	28	17	4	1	
Anger				1	19	31	1	30	22	2
Suffering				1	6	14	70	3	2	1
Determination		1			1	7		52		2
Disgust					1	1	2	1	63	7
Contempt						2		3	12	86
Scattering	12/100	6/100	/100	10/100	2/100	9/100	7/100	4/100	/100	2/100

19 poses, two for each of the named emotions except Determination, with judgment of each pose by 100 observers. The entries give frequencies in per cent; the "scattering" judgments could not be located on the scale.

> The table undoubtedly shows agreement between the judges rather than unlimited scatter. It shows a positive correlation between pose and judgment. However, the attempt to obtain a linear scale has not been wholly successful. 'Suffering' and 'determination' cannot be perfectly fitted into the series. According to the frequencies, suffering belongs next to fear, and determination next to anger; but fear and anger, still according to the frequencies, must not be separated. If suffering were located on a branch line from fear, and determination on a branch line from anger, the main line would be quite regular.
>
> The first three emotions in the series, love, happiness, and mirth, overlap greatly with each other but scarcely at all with any of the others. If these three are combined, and if suffering is combined with fear, and determination with anger, we have a scale of 6 steps which is very satisfactory as far as the frequencies go. The judgment is seldom displaced more than one step in this coarse scale, and if the steps are assumed to be equal, the correlation of pose and judgment is + .92. Other data, from Ruckmick (1921, and from Gates (1923), obtained with other sets of poses, fit neatly into the same scale. (p. 250-251, 53)

Woodworth seems to have had an intuitive basis for preliminary selection of categories and their arrangement. Perhaps this may have been modified somewhat by the same order being imposed on the poses and the desire to accommodate modal categories close to other categories with high frequency of response, in a "best fit." All this is speculation, though, and no explicit or systematic procedure seems to have been employed. This is a very serious defect, as we do not know if other experimenters could arrive at the same scale or even a synonymous one. We do not even know the effect of his arbitrary scale assumptions, as to extremes, units, or what effects the presence of a "scattering" category has on his scale.

But, on the whole let us suppose that reasonable criteria could be formulated, the satisfaction of which would enable any experimenter to derive the same scale from the same data.

There is an important unresolved question here: How many of the properties of the ensuing scale are empirically based and how many are a resultant of the conditions the experimenter has imposed upon the data? In order that the results may assume a form more meaningful to him, has the experimenter put the data into a form which predicates at least part of his results? Is it likely that a form derived to be amenable to one type of theoretical treatment will be equally well suited to other systems? The possibility of refutability is vital especially if we hope to use such data to discriminate between theoretical systems by confirming or disconfirming hypotheses derived from these systems. Since a major function of empirical observations is precisely in deciding between alternative predictions derived from competing theories (and perhaps even in all normal human perception), we lose a major portion of its value when our data contain an implicit theoretical bias. If in the use of the data, we forget the source of those non-empirical considerations which guided its formulation, we risk the danger of confirming a proposition with data which, in more neutral formulation, would more likely confirm a hypothesis logically mutually inconsistent. A thorough mixture of empiricism and a priori explanation cannot satisfactorily perform the function of either. Either principle is meaningless if not accessible to future modification or inductive generalization, and is useless as a basis for further scientific procedures of discovery.

This importance of empirical interpretation is further augmented when, as is frequently the case, unrealistic constraints are imposed upon the perceiver. To date, too little concern has been directed towards discovering the actual process of perceptual discrimination which may be

utilized in ordinary perception of emotion. Such studies have dealt primarily with verbally denoted categories, assuming that individuals customarily (or could as well) class their perceptions into one of a pre-selected set of categories which would be the same for a majority of perceivers. Further, one must assume they are equally disposed to use the same label, under which the perceptions were originally classed, to describe (a quite different function) them when requested at a later date under different conditions (standardized in experimental situation).

The psychological set to classify and the set to describe vary considerably in their respective goals, the means utilized to attain these goals, the criteria (formal or social) they must satisfy, and the freedom of expression allowed. It is unrealistic to assume any greater precision and uniformity of usage under these disparate conditions than exist in everyday usage (where precision and uniformity is uncommon).

Secondly, the assumption of constancy over time is quite unwarranted. Such a presumed independence of memory from reorganizational thought factors and motivational influence is not justifiable, even when the same type of response is specified. Such influence may even heighten the disparity when different response forms are used - spontaneous, classification or recall.

Extreme care is necessary to insure that our investigation bears some relation to the actual process of the perception of emotion. This must become one of our primary goals in any future refinement.

Returning to the mode of combining the data itself, the experimenter's choice of a mode of analysis greatly affects the inferences we can make about the data. The very difficulty of ordering similarity by successive

degrees of overlap led to the refinement whereby quantified data were summed across several perceivers into a "mean social position."

Schlosberg (42) assigned a number to each category on Woodworth's scale, crediting each picture with a certain number every time it was classed into the corresponding category. By summation and division he was able to compute the arithmetic mean scale position for each picture on the Woodworth scale. The average deviation was also computed. This could be used as a symmetric weighted measure of overlap and presumably could enable one to arrive at a more powerful scale. But this is possible only after the order of the categories has been determined or assumed - in this case by the order of the natural integral numbers assigned to each category. However,this assignment, as far as can be determined, is still governed largely by the experimenter's intuition, in a trial and error process of fitting.

Indeed, Schlosberg (42) noticed a high degree of overlap between categories at either end of the numerical scale. He first hypothesized that the scale was in fact, circular, with the categories arranged on the scale along the circumference. Guided by analogy with the color circle, he then imposed a set of two dimensions upon the plan of the circular scale. In another experiment (43) he had judges assign numbers to Frois-Wittman expressions on each of these dimensions (a rating scale from 1 to 9). The ratings were averaged for each expression and plotted in this two dimensional space. Each expression was thus characterized as a point in this two dimensional space. A vector was drawn from the origin (representing the ideal midpoint of both scales, a rating of 5 on each) through each point to the circumference, thus determining a corresponding position

on the circular scale. These "predicted" scale values were later compared with the values obtained by sorting stimuli into categories and computing mean scale positions. Very high agreement was obtained between the two measures correlated on the poses(21). Correlations greater than .90 were obtained when the same procedure was tried on other series of facial expressions. This demonstrates that rating on component scales can with reasonable accuracy predict the rating derived from category assignment measures, although the reverse need not hold. That one can predict from behavior guided by one form of verbal constraint, behavior guided by another form of verbal constraint is not, in itself, particularly surprising.

Examining the two experimental situations from the standpoint of how well the response potential "matches" the stimulus potential we find considerable differences. In the category situation, the perceiver can have, at the time of judging, an almost free range of possibly relevant attributes although available categories probably exert a suggestive effect. However, the responses are sharply constrained as they may consist only of classification into exactly one category and non-classification into the other categories specified. In the second situation, the relevant attributes have, presumably, been restricted while considerable flexibility of response has been accorded. In the category situation, the greatest restriction is imposed on the response side, whereas in the component situation, the greatest restriction is imposed on the stimulus side. In either case, the validity of the results for the study of the perception of emotion is adversely affected. It is worthwhile to note that whatever the nature of the restricting force may be, it seems to produce much the same effect when employed either on the stimulus or the response side.

In order to infer positions on the category scale, Schlosberg assumed that each category was represented by an equal unit on the scale, regardless of how often it was used. Presumably, in order to arrange the categories, only the average deviation of an expression on the scale was used as a measure of overlap. This requires certain assumptions and restricts its meaningfulness as to the nature of underlying distribution. To make his predictions, Schlosberg assumed orthogonality of the dimensions he specified. Their complete independence of each other is, however, an empirical matter, and cannot be decided by fiat. Their assumed independence would undergo question on intuitive grounds alone.

These studies are valuable chiefly for indicating an escape from the dilemma posed by using categories and their attendant inaccuracies. Thus, Woodworth analyzed the errors and used the magnitude of discrimination failure to construct a scale, perhaps the first explicit recognition that basic dimensions may be of more import in judging facial expressions than verbalized categories. Schlosberg seems to be the first experimenter since Wundt (55), or Duffy (16) to emphasize that several factors may be interacting simultaneously in emotional processes when he hypothesized two (later extended to three) (44) dimensions of emotion in facial expression data.

Such methods of analysis already contain implicit theoretical assumptions as to which conceptions of emotion are important in discrimination (as any a priori categorization prejudges the way the world of emotion may be sliced). As such, they do not constitute a truly empirical attempt to find those dimensions one normally uses in perceiving

emotional behavior. Considering the extreme suggestibility of the perceiver with respect to what verbal limitations are imposed, which has been demonstrated in previous data, this entails no necessity that imposed dimensions have any real existence outside of the test situation, in spite of how reliable rating behavior may seem within the situation. In fact, Munn (36) has found that allowing perceivers to use their own categories results in greater agreement on categories than occurs if perceivers are constrained to use a standard set! Preferably, one should use only as many dimensions as are needed to explain the data, rather than impose all the data on a preselected set, whether or not it fits. In the latter case, resultant error is unanalyzable without making severe assumptions about the underlying distribution of responses.

Furthermore, in studies such as Schlosberg's (43), the individual is constrained to use each dimension separately. Such artificial separation ignores the complex relational structure between emotions and dimensions. In addition, no acceptable method of assessing the relative importance of dimensions in the perception of emotion from facial expressions is available. The result that facial expressions may be rated reliably on these particular dimensions is not significant for a descriptive systematization of emotion unless one can demonstrate that these dimensions are ones which are likely to be used in perception of emotion, or ones which should be used, either for the sake of logical clarity or conformity with empirical data. Thus, perception of facial expressions has not been shown to be conventional, it has only been required, as a step in analysis, to be so.

From a scaling or measurement viewpoint, there remains much to be desired in construction of the foregoing scales. Several types of scales

of measurement may be distinguished: nominal, ordinal, interval or ratio - corresponding to the procedure of analysis employed and in increasing order of the number of properties assumed (50). It is useful at this stage to see what types of scales are being attempted, since these scales constitute the meaningful form into which raw data will be channeled so that it may become systematically communicable.

If one regards scaling as a process of linking a formal descriptive model to some discriminable aspect of objects or events, then it is possible to conceive of a scale of measurement, in general, as a set of rules assigning numbers to objects or events (13). A scale of measurement is usually presumed, i.e., one often assumes that an isomorphism exists between properties of a set of numbers and the empirical operations one can perform with aspects of objects or events. "This isomorphism, of course, is only partial. Not all the properties of number and not all the properties of objects can be paired off in a systematic correspondence. But some properties of objects can be related by semantical rules to properties of the numeral series" (50).

This is precisely what is misleading as few experimenters attempt to ascertain which object properties may be validly related, much less try to ascertain the properties of the number set selected. Characteristically the procedure most often employed is to assign rational numbers (usually positive integers) to any conceivable property in which interested, and thereafter to regard any operation applicable to real numbers as employable. The difficulty lies in the interpretation of the ensuing results.

Since a primary purpose of mapping responses into a formal systematic descriptive model is to substitute operations in the abstract system

for operations on the object system, we can scarcely expect that use of operations for which no empirical analogue can be found will insure a resultant which reflects any "empirical object," or has any empirical meaning. That is, in constructing a descriptive model, one must take care that its formal axioms are satisfied, both by mapping object aspects and empirical operations on the objects' aspects onto the set of all primitives within the system. (The surreptious addition of formal operations with no empirical correspondents cannot be tolerated unless we are prepared to substitute hypotheses for descriptive goals.)

In this respect, Woodworth attempted to construct an ordinal scale from his nominal scale data with the aid of indices of dispersion. Schlosberg attempted the construction of a ratio scale directly from nominal scale data, while we yet have no idea what the addition of such stringent scale assumptions involve in terms of permissible inferences about emotion. Our purpose in constructing a descriptive model is subverted if there is no guarantee that permissible operations on the data have no empirical significance. Schlosberg assumes that the intervals are <u>equal-appearing</u>, assuming a unit of measurement and a particular distribution. Indeed, he must assume this, since the means and average deviations which he uses to "order" the categories are in error to the extent that the successive intervals on the scale are unequal in size. Unfortunately, however, this has not been the case in data summarized in reports of his experiments. Even though the data obtained does not bear out this assumption, there remains the possibility, however obscure, that the population of responses which he is sampling may be distributed in a rectangular fashion. However, there is very

little evidence to support such a contention. The psychological responses which have been investigated extensively have often tended to assume a Gaussian or normal distribution. Indeed, most of the statistical indices in common use presume it. In the absence of supporting evidence from either neighboring areas or in samples of the response itself, a rectangular distribution seems unrealistic to assume. Consequently, means, average deviations, and linear interpolation (to determine the midpoint of a class interval) have little empirical meaning raising serious questions about the ordering process itself. Uncertainty as to how much of the ordinal scale is determined by the phenomenon being observed and how much by the method of analysis imposed is undesirable for a basis for scientific description.

Beyond the ordinal scale assumptions, even greater difficulties arise in the construction of interval or ratio scales. Here one is forced to postulate operations which as yet have no empirical correspondents. "An interval scale can be erected only provided we have an operation for determining equality of intervals, for determining greater or less, and for determining equality (not greater and not less). To these operations must be added a method for ascertaining equality of ratios if a ratio scale is to be achieved" (59). In general this takes the form of the problem of selecting an invariant "unit" of measurement and, for ratio scales, a "zero point." If and when these operations are satisfied, one can proceed to construct a ratio scale and employ the usual statistical measures, assuming an underlying normal distribution of responses.

Unfortunately Schlosberg has not demonstrated that these operations have any empirical counterpart or meaning. The continued postulation of

unobservables should not be encouraged in descriptive data and could conceivably distort the data without one knowing how. This type of procedure may be responsible for the anomalous result that the stimuli seem to cluster considerably above the assumed midpoint on one of Schlosberg's dimensions. Even if one could avoid the above problems of imbuing the assumed operations with checkable meanings in terms of responses, and ignore the numerous, inconsistent distribution assumptions necessary to employ these assumed operations, one is left in a void as to their possible significance in terms of emotion itself. This is of greatest concern when interpretations of the results are attempted.

Adherents to the "fruitfulness" view of science may claim that it does no harm to obtain results by mathematical or other symbolic form of manipulation to which no present observable corresponds. After all, many subatomic particles "existed" first only in mathematical form, receiving empirical corroboration only later. This view is not without merit with regard to the usefulness of competing theories (although it is by no means invulnerable to critique), but should involve a minimum of theoretical adjudication. When, because of our particular "choice" of a procedure of data analysis, one arrives at a resultant with no observable counterpart and no unravelable relation to any other theoretical or empirical procedure, one has progressed little towards the provision and establishment of a universal basis for scientific description.

Even the nature of the responses imposed on the perceiver can severely influence the nature of our results. For example, it is questionable whether it is meaningful for either experimenter or perceiver to assign certain numbers to states which apparently cannot even be reliably

characterized by names frequently associated with them. The postulation of extraneous and possibly meaningless entities and operations can confuse the situation when we attempt to find a real world interpretation satisfying all the primitives in such a formal system. In order to insure that a system has generality and empirical meaning, every primitive in our system should be required to indicate the possibility of its own verification. Then when the user wishes to speculate or specialize, he can make appropriate modifications in his system and be aware at just what point they are necessitated or justified.

If in any sense one hopes to discover the important factors and their interrelations, one must not prejudge the issue by preselection of categories which one is compelled to use. What would happen then if one avoids forcing the perception of expressions of emotion into pre-set language categories. Just such a procedure was employed in a preliminary study to be described in the next chapter.

Chapter III

PRELIMINARY EXPERIMENTS

As a necessary preliminary to the main investigation, several informal experiments were conducted. A new technique of data collection and analysis was attempted in the first experiment. The other preliminary investigations tried out different designs of stimulus faces.

The First Experiment

The first preliminary experiment to be discussed required perceivers to judge with a minimum of language constraints the relative similarity of samples of behavior. The analysis of data was an attempt to determine empirically a set of dimensions which would characterize such judgments on a selected subset of the pictures used by Schlosberg (54). Eight Frois-Wittman pictures of facial expression were used as stimuli, presented three at a time at the corners of an equilateral triangle projected on a screen. These particular stimuli were chosen as representing maximal mutual distance from each other in terms of Schlosberg's two-dimensional circular scale diagram(54). There were 56 presentations in all, representing all possible combinations of the stimuli, taken three at a time. Subjects were instructed to judge at each presentation which two pictures were most alike and which two were least alike with respect to the feeling-state of the individual pictured. These relative similarity comparisons were regarded as determining an

ordering of Euclidean distances in the psychological "similarity-space" of the individual. That is, when a pair of stimuli in a triad is judged least alike, the psychological distance between them is inferred to be greatest among the three stimulus distances. The non-metric technique of analysis employed, hereafter called the Multidimensional Unfolding Technique, is discussed in detail in Chapter V. Essentially, order "dimensions" may be constructed by taking the longest distance and fitting as many as possible of the distance comparisons along it. This process is repeated until all distance comparisons "project" on at least one dimension. In this particular experiment, three dimensions sufficed to account for the judgments given.

This analysis was performed on each subject's set of judgments and for the group of all 17 subjects (analyzing the triads constructed for the group as a whole). By informal inspection, individual scales did not differ substantially from the group scales.

The main features of this method are its multidimensional and non-metric character. Minimal constraints are placed either by the method of data collection or data analysis upon the number or location of dimensions obtainable, and no unit of measurement or zero point must be specified. It would seem that the distance function postulated need not even be Euclidean, and generalization to other types of distance functions is possible. The only requirement is that certain conditions must be satisfied before one can construct order dimensions from the data. That is, the term "dimension" is used here in a sense analogous to a principal factor in factor analysis. More formally, a dimension corresponds to a representative set of elements, between any pair of which it is possible to construct, from at least a connected subset of the data, a transitive, antisymmetric relation.

Considering the great degree of freedom allowed the judges, who were uninstructed about any attributes, and the radically different procedure of analysis employed, it was somewhat surprising to obtain dimensions which correspond in order of stimuli very closely with Schlosberg's three a priori dimensions (Schlosberg added a third dimension at a later date) (44). Yet the status of these dimensions is still largely undetermined since his procedure can hardly be appealed to as a validation. Even within his system the status of the third dimension (which he calls Tension-Sleep) is ambiguous. The basis for its addition is not entirely clear - probably partaking much of the recent popularity of measures of the level of activation (35) occasioned by the recent evidence (15) of the influence of the reticular formation on neural activity. A further clue to its arbitrary nature is its apparent monopolarity. Schlosberg attempts to account for the relative absence of stimuli at the Sleep end of the scale by observing that <u>emotional</u> expressions are likely to be predominantly at the Tension end of the scale. Certainly no published study of facial expressions of emotion has attempted to control for this factor. The absence of any objectively formulated rationale for the selection of expressions for analysis is a major consideration in favor of selecting facial expressions that may be systematically varied.

Although the response data in this preliminary experiment is collected in a manner which obviates the necessity for a priori verbal categorization and is analyzed into forms utilizing a minimal number of theoretical assumptions, the problem of a more concrete identification of these dimensions remains to be solved. Furthermore, it can be

seen that the stimulus aspects - the facial expression stimuli themselves - have been investigated in a rather haphazard manner with little thought as to their comparability or selection. The treatment of these unsolved problems, then, must constitute the core of any detailed empirical investigation of the perception of emotion.

The latter problem, concerning the selection and comparability of stimuli, is the subject of the second preliminary investigation.

The Second Experiment

Several difficulties discussed in Chapter III led to the consideration of the use of schematic faces as stimuli. The avoidance of certain interactions, whether based on assumed similarity or on idiosyncratic reactions, between the perceiver and the perceptual objects must be a primary consideration. The maintenance of comparability across stimuli is difficult with any but schematic stimuli. The standardization of schematic faces insures equivalent perceptual configurations for all perceivers, since selective perceptual sampling is more apt to focus on potent physiognomic areas. The faces need not be so complex that one is forced to attend to only a small sample of those characteristics present in the stimuli. Standardization further allows the experimenter to identify the cues which mediate the perceptual judgments. A certain degree of experimental control over the stimuli is thus afforded and the procedure of sampling employed need not be haphazard and unrepeatable. Finally, use of schematic faces aids in avoiding the problem of inaccuracy.

When the entire set of stimuli is internally comparable in terms of stimulus characteristics per se, one can detour the problem of individual validity on which endless research has stumbled, since here the problem

is not what particular expressions may convey, but what are the properties of the entire set of facial expressions. Certainly, the introduction of schematic faces avoids this question in that we do not inquire (ordinarily) "What does he really feel, or what is he really like?" of a schematic face. The interest here is in what kinds of ranges of emotion are conceivable in facial expression; thus, facial expression is used more as a set of stimuli which excite a response process, rather than as a process in itself. Perceptual impressions are called out by this impoverished situation, but they are not so determined in form that they must be conventionalized by the individual judge by all the cultural constraints, including evaluation of accuracy, it is possible to muster.

As noted above in the first preliminary study, an attempt was made to select stimuli that were maximally different from each other as measured on Schlosberg's two dimensional circular scale diagram. However, the addition of more stimuli and dimensions in the absence of comparability of cues between stimuli renders this concept of discriminability ambiguous and ungeneralizeable.

There are several reasons for devoting so much attention to sampling. Primarily, of course, it is desirable that relations holding within the data sample can be generalized to a wide number of situations. In this sense, a general scheme is superior to specific faces which often convey specific situational impressions.

Acting on the information about schematic faces presented in Chapter I, facial schemas were constructed with mobile cues systematically varied. This second study was conducted by constructing schemas employing all possible cue variations and presenting the schemas for rating by judges

on several different scales, including a judgment of "normality." A summary of the most significant results follows. Mouth height, distance between eyes, and nose length appeared irrelevant for the study of emotion, presumably because they are constant for a given individual. Where a choice of values was necessitated in choosing cues to be included in the set of faces, the most neutral or "normal" looking value was employed.

Mouth length and facial tension curves were also found to be rather irrelevant or inconsistent in inference about emotion. Eyebrow shape lent itself to a considerable number of idiosyncratic judgments. An informal glance at a large number of photographs reveals little standardization of eyebrow shape or kinds of shapes. The case is similar for mouth and eye shapes, so no attempt was made at simulation of average population shapes. Smooth curves (ellipses or arcs of ellipses) were employed with the central value being fixed at the most "normal"-looking value. Certainly the psychological neutrality here is more vital than any physical neutrality - which is always relative to other physical factors which can be measured.

In keeping with this standardization of reproducible forms which otherwise would be difficult to reduce to a populational average of physical parameters, a circular facial outline was used to enclose the facial expression. Brunswik (7) used a modified oval (egg-shaped) face. The present study initially attempted use of oval shapes, which, however, occasioned much comment of an idiosyncratic nature from the perceivers. Facial shape is a predominantly structural feature and as such should be irrelevant to a study of facial expression. Any attempt at simulation of a particular facial outline can only be a distorting influence.

Facial expressions were then schematized without the aid of enclosure. This procedure rendered exceedingly difficult the already rigorous task, so the "disembodied" facial expression was rejected in favor of the circular outline whose symbolic value was relatively uniform for all perceivers. A further concession to simplicity was the use of a symmetric form. That is, all faces were designed to be symmetric about the nasal line, features on one side were in the same relative position as features on the other side. Although it is not safe to assume that human facial expressions are completely symmetric, there is a physiological tendency for spontaneous facial expression to be symmetric and therefore it seems to be the safest assumption under present circumstances.

The final form employed was enclosed by a circular outline. There were three variable features, mouth curvature, eye aperture and height of eyebrow over the eye, with three possible values for each elliptical form; there were also two constant features, a dot for the nose and a dot for each eye.

All possible combinations of these stimulus variations were formed, 27 in all. Each face was rated from 1 to 5 on several scales. These scales were suggested either by the individual subjects themselves, or by previous experimenters and theorists (44, 49). Greatest attention was paid to those scales for which the greatest empirical evidence existed. The final selection of maximally discriminative stimuli should include stimuli which are distinctive on at least one of these scales. For example, scales of Pleasantness-Unpleasantness, Attention-Rejection, Tension-Sleep, Approaching-Avoiding, Organized-Disorganized, Aware-Unaware, were extensively investigated.

Even this limited selection of scales or psychological attributes often proved difficult to apply. That is, many subjects could not understand the terminology, found the terms difficult to apply to emotion or its expression, or could not accept the scales as bipolar. An example of the kind of results obtained is presented in Appendix A. In this informal table, the abbreviations at the left signify the particular configurations of facial cues. Each configuration or face was rated on each of the scales listed, by at least 35 subjects. The mean ratings of each face on each scale it was judged are given in the table. On the basis of tables like this one, particular faces were selected and redrawn on a larger scale. The mean ratings presented in this table should be interpreted with caution, as the presentation of rating scales was not random and the configurations of cues were somewhat crude and differed in the number of possible cue variations.

A total of eight faces was selected from the 27 so that comparison with previous experiments with human faces could be effected. Those faces which received the most extreme average ratings on each scale were selected. Since some faces occupied extremes on more than one scale final selection of the set of eight was also based on a consideration of maximal distance. That is, if a stimulus face is not at an extreme on a particular scale, it should occupy a central value. Each cue variation was represented an approximately equal number of times in the final set of eight stimulus faces. These eight faces which are maximally discriminated, in some sense of the word, were then employed as stimuli in the main experiment, described in Chapter V.

Chapter IV

THE SPECIFIC PROBLEM

Introduction

At this juncture, it is well to review the objectives and summarize the questions which the use of the present type of procedure may raise and therefore the kinds of results which may be anticipated.

The problem under study is to find, then, a characteristic set of dimensions or psychological attributes which best describe observed variations in facial expressions (presumed to be the major channel by which emotions are ordinarily expressed). For this, one needs:

1. A method for obtaining judgments of expressions which does not necessitate the use of categories or numbers.
2. A method for analyzing the response data obtained to recover the relations between the stimuli.
3. A method for identifying or labelling these relations or dimensions.
4. A way to identify those physiognomic characteristics which contribute to facial expressions of emotion.
5. Some rationale for the selection of stimuli.
6. Assessment of the effect of constraining information as produced by variations in instructional set.

For 1, the use of the Method of Similarities (12) is indicated. This method does require a response which forces the perceiver to consider similarity or dissimilarity between each two stimuli, irrespective

of whether such relations would ordinarily be considered outside of the experimental procedure. More serious is the requirement that if triads of stimuli are to be judged, the three similarity judgments must be ordered in magnitude. Although such a requirement may not violate any phenomenal relationships, this condition of transitivity within a triad is probably somewhat strong to require when the individual is inferring emotional characteristics. Intransitive judgments could arise from mutually exclusive "logic-tight compartments." Nevertheless certain assumptions must be made before any analysis of judgmental data can begin and this assumption of transitivity seems to be the most justifiable, at least within a single presentation.

For 2, use of the Multidimensional Unfolding Technique permits incompatibility between different triads (in the sense that one dimension need not represent the order relations given in all triad judgments) and utilizes it for reconstruction of order dimensions. These order dimensions can serve as basis for the empirical construction of more powerful scales of measurement. The relations between stimuli can be circumscribed, although the exact nature of a relation beyond its order properties cannot be identified nor spelled out in detail until more powerful conditions are imposed. As in all experiments, the behavior of the subject is not expected to undergo drastic modification over the response period. To insure that responses made during a particular time interval are representative of the entire response period, a procedure for the presentation of stimuli in a randomized sequence must be adopted. In a study of emotion, additional steps should be taken to minimize the possibility of mood shifts in the subjects.

For 3 and 6, one must effect a translation of phenomenological concepts into verbal concepts and vice versa. Of course, this is not automatically an isomorphism as is customarily assumed. Efforts will be made to identify and label the dimensions used by incorporating within the experiment itself variations in instructional set to some of the subjects.

For 4, one can expect that the employment of schematic faces will greatly simplify the identification of relevant physiognomic characteristics. But one cannot expect a simple monotone relationship between variations on a single cue and the dimensions obtained, as undoubtedly the total pattern of cues, even within a single face, form a "Gestalt" of their own.

For 5, the rationale for selection discussed in connection with studies one and two in the preceding chapter is directed towards the sampling of maximally discriminable stimuli. Less discriminable stimuli, even subliminal cues, may bear much greater responsibility in judging emotion than is assumed by this sampling procedure. Although these cues may mediate the judgments of similarity, there is no guarantee that the set of dimensions or psychological attributes which characterize stimuli of great discriminability are the same as the relations between less discriminable stimuli. Thus any set of dimensions which may be discovered cannot be considered as exhaustive, but merely as basic dimensions which delineate and characterize the most discriminable aspects of expression.

Hypotheses

Because of earlier research on human facial expressions, there are several characteristics of the set of perceptual dimensions which

one may look for. Although the very nature of the problem proposed is to discourage any premature hypothesizing, certain expectations may be formulated. Since no other satisfactory referential basis exists, the endeavor to find a basis for the description of emotion makes it misleading to express these expectations in the form of precise hypotheses.

1. In judging schematic faces, the perceiver will use three dimensions. Three dimensions have been found in the perception of emotion from human faces, using either Schlosberg's rating and averaging method or the present Method of Similarities and Multidimensional Unfolding Technique.

2. The dimensions utilized in the judgment of emotion from schematic expressions will be the same basic ones as those in perception of emotion from human expression.

It is reasonable to expect configurations of schematic faces similar to ones obtained with a procedure differing only in the stimuli utilized. Hypothetically, though, one might expect more dimensions to be utilized as the complexity of the stimuli is increased by the use of actual human faces. This effect may be neutralized, however, by the canalizing effect of the perceiver's own selective sampling of human facial characteristics.

3. The order of saliency of dimensions in judging schematic faces will be the same as when judging human faces.

Since the order of saliency or importance is determined chiefly by the order in which each dimension is uncovered in the Multidimensional Unfolding Technique, a procedure analogous to the method of analysis by principal components in factor analysis, a change in stimuli should have no effect on the order unless a change in their conception is occasioned.

Because evaluative concepts so often permeate judgments, one could expect that a dimension with some evaluation of "Pleasantness" (perhaps the easiest to detect or the most salient of characteristics of emotional behavior) would be primary. This dimension was primary in each of the two investigations on human faces to date. Mouth curvature is probably the primary mediating cue.

Additional dimensions may be less salient but should appear in the same order as in the two studies on human faces - "Attention-Rejection," then "Tension-Sleep." Their mediation can only be guessed at. Eye aperture has been suggested as a cue for either, in combination with certain other configurations of characteristics.

4. A perceiver instructed to judge on a particular attribute will use primarily the corresponding dimension. That is, the primary dimension obtained in the analysis of the perceiver's response data will represent the attribute previously specified to him. This is a procedure for identifying the dimensions used in "spontaneous" perception, that is, when the judge is uninstructed on any particular attribute.

5. A subject instructed to judge on a particular attribute will use primarily the same dimension as a subject instructed to judge on the attribute's "logical opposite."

This hypothesis, if corroborated, would enable us to construct bipolar dimensions rather than assume them a priori. Much confusion has arisen in the past when experimenters have attempted to define dimensions with concepts which are neither logically nor experimentally bipolar.

Chapter V

PROCEDURES AND RESULTS

The present experiment can be divided into three components, each a separate study utilizing different sets of experimental subjects and different procedures.

STUDY I

Purpose

This study constitutes the basic effort to discover what dimensions are used in the perception of emotion from schematic facial expressions, with no particular attribute specified.

Subjects

The subjects were 13 (8 male, 5 female) students taking beginning psychology who had not previously served in a psychological experiment. They were selected from the University of Michigan Psychology Department subject pool. Their ages ranged from 17 to 26, with a median age of 18.

Procedure

The procedure followed was approximately the same as in the preliminary study of facial expressions, described in Chapter IV.

Upon entering the experimental room, subjects were seated in alternate seats and given a response form. The experimenter read a statement (Appendix B) designed to arouse interest and elicit cooperation in the experiment. The subjects read the instruction sheet (Appendix C) and any questions were answered.

The stimuli presented were the final 8 schematic faces selected in Chapter IV. These schematic faces are reproduced in Appendix F. They were presented three at a time at the corners of an equilateral triangle projected on a screen, at approximately 30 second intervals in a randomized sequence. Although the subjects were advised to proceed quickly, there was no actual time limit. The presentations were preceded by 10 minutes for instruction. The entire session was 45 minutes in duration. There were 56 presentations in all -- that is, all possible combinations of 8 stimuli, 3 at a time. Subjects were instructed to judge at each presentation which two faces were most alike and which two were least alike, in respect to the feeling state depicted.

Method of Analysis

The responses were converted into a form, suggested by Hays (24), which is more amenable to analysis by regarding judged dissimilarity as a distance. That is, when a pair of stimuli in a triad is judged least alike, we infer that the distance between them is greatest. For each subject, each triad judgment was converted into a four place symbol as follows. Suppose the stimuli in a particular triad are denoted by B, C, and F. If B and F are judged as most alike and C and F are judged as least alike, this can be denoted by the symbol C BF.

Each individual triad judgment can be converted into a group triad by examining the particular triad for each subject. That is, select as the greatest distance in the group triad the pair of stimuli which is most frequently the farthest apart in that particular triad, across all subjects. Similarly, that pair which is most frequently closest together is selected as the smallest distance in the group triad.

All triads in the set judged can be converted into a group form in this manner. This group form should minimize the error due to mistakes in responding, lapses in attention, extreme deviancy or uncooperativeness. Since the experimental session is quite exhausting for the subject and the Multidimensional Unfolding Technique is sensitive to small variations in response, the risk of incomplete data is great and suggests the use of replications over individuals.

The group triads then can be converted into a characteristic set of dimensions by means of the Multidimensional Unfolding Technique. A full description of this procedure may be found in Hays (25). To summarize,

> The possibility of a non-metric analysis of similarities judgments was first suggested by Coombs, who proposed the Unfolding Technique as a method for analyzing such judgments more or less directly, without the necessity for ad hoc assumptions about distributions, etc. The Unfolding Technique is applicable only to the case in which one and only one attribute is the basis for the judgments of the subject. It deals with the rank orders of decreasing similarity (I-scales) of the remaining stimuli from each stimulus in turn. By regarding these I-scales as an order of distance from a point on a line to each of the other N - 1 other points on the line, one is able to infer the order of the stimulus points on the attribute (line) which was the basis of judgment. In addition, one is able to infer the order of magnitude of at least some of the distances between the stimulus points in this way. However, the Unfolding Technique is applicable only when the I-scales satisfy certain properties indicating that there is actually only one latent attribute underlying the judged similarity of the stimuli.
>
> The method of analysis for similarity data introduced here is an extension of the Unfolding Technique to the case in which the attribute space underlying the data is multidimensional. The procedure starts with the same assumption as Coombs': dissimilarity is equivalent to distance between two points in one or more dimensions. Then by geometrical considerations which seem to be true for points in some arbitrary dimensionality r, it is possible to recover one

> or more rank order 'scales' or 'dimensions' from a set of N I-scales. Each of these rank-order scales represents an order of projections upon a reference axis in the attribute space common to all of the stimuli. These reference axes are chosen as a 'best fit' to the relative positions of the stimulus points in the space in a 'least squares' sense. (pp. 2-3, 25)

A mathematical rationale is then presented.

> In short, the principle involved here is simply that in order to minimize residual for a trio of points in terms of their projections upon an axis, a necessary condition consists of making that scalar product corresponding to the largest angle of the configuration negative. In even simpler terms, one must choose the axis in such a way that the pair of points having the most distant projections upon the axis are those points which are the most distant in the configuration of three points. In the remainder of this discussion this principle will be referred to as the 'betweenness condition.'
>
> The major thread of this discussion is, then, that using ordinal properties in the data alone, we are able to ascertain within rank order a set of reference axes satisfying at least a necessary condition for a least squares fit to the points in the attribute space underlying the similarity judgments. The idea of the solution becomes simply to start with the longest line in the configuration, and successively, for each triad of points, to insure that the longest line in the triad has the largest projected distance upon the reference axis, as shown by the rank order of the three points involved. Then having arrived at an initial rank order which fits this betweenness requirement optimally, we construct a second rank order satisfying the betweenness requirement optimally for those triads which were not satisfied by the first rank order, and which duplicates as few of the triad rank orders of the first reference axis as possible. This triad 'fitting' process is carried out until each triad fulfills the betweenness condition for at least one reference axis. (pp. 9-11, 25)

Briefly, in practice, the entire set of triads is examined and the distance which is most frequently chosen longest is selected as defining the first dimension. A dimension so chosen certainly accommodates a maximal number of "relevant" triads (all those triads in which the two dimension endpoints appear).

The two I-scales between the two stimuli may then be examined. The order of decreasing similarity to one stimulus should correspond roughly to the order of increasing similarity to the other stimulus. In practice, complete "mirror-image" I scales are relatively infrequent. Often, pairs of stimuli are "reversed" in order with respect to the rest of the scale. On such occasions, one may examine the especially relevant triads (those triads composed entirely of the two non-reversing stimuli and the two endpoints of the dimension) for further indications of order. These methods are merely aids, though - the ultimate criterion for construction of an order on a dimension is the attainment of the greatest number of triad projections or triads satisfying a "betweenness condition," on the dimension. This process may be repeated with the remaining unsatisfied triads until each triad projects on at least one dimension. To insure their relative independence, each dimension is constructed so as to duplicate as few as possible of the triad rank orders of the previously chosen reference axes. Thus a quasi-independent set of dimensions can be constructed to characterize the set of judgments of emotion from schematic facial expressions.

Results for Study I

The results of Study I, in the form of dimensions obtained, are presented in Table 2. These dimensions are portrayed in Appendix F. Three dimensions were obtained, accounting for all the triad's distance relations. While this is the same number as was obtained in a similar study on human faces, there are a fewer number of triads (in this case, three, while with human faces, seven) responsible for our inference of an order on a third dimension.

Table 2.

General Dimensions of Judgment of Emotion from Schematic Facial Expressions

	Triad Projections	
	Initial*	Final**
D_1 E B H F C D A G	39	39
D_2 B A E C F H G D	14	28
D_3 C A G B E D F H	3	16

	Relative Independence of Dimensions***	
	X	τ
D_1 with D_2	10	.29
D_1 with D_3	10	.29
D_2 with D_3	10	.29

*Initial triad projections are those orders which each dimension is chosen to "describe."

**Final triad projections are those orders ultimately "described" by each dimension.

***X and τ (tau) are measures of independence described in Study II.

Intuitive examination of the physical characteristics of the stimuli on the first dimension reveals down-turned mouths at one end of the scale, up-turned mouths at the other end, with straight mouths in between. The other physical cues show no systematic relationship, although eye aperture tends to be medial with low eyebrow height, at the down-turned mouth end of the scale. An intuitive conclusion would be that D_1 represents some sort of "happiness" attribute mediated primarily by mouth curvature and displaying a tendency for medium eye aperture and low eyebrow height at the "sad" end of the dimension.

The second dimension seems to be of a more hybrid nature. There is again the distribution, with one exception, of upturned mouths at one end, down-turned mouths at the other end, straight mouths in the middle. This time, however, all three faces with wide eye aperture are at the upturned mouth end (on D_1, the faces with wide eye aperture were more widely distributed). Eyebrow height is unsystematically distributed, with perhaps a slight tendency for medial cues to be located at the down-turned mouth end of the dimension. Intuitively, this dimension seems highly similar to the first dimension. The only major difference seems to be the prevalence of faces with wide aperture eyes at the upturned mouth end of the scale. This seems consistent with the hypothesis that this dimension is related to attention or attributes characterized by impressions of high receptivity on the faces which may be seen as actively approaching external stimuli. Such is the description by Schlosberg of his own second dimension.

The third dimension shows some surprising results. This time, straight mouths seem to be distributed at both extremes, with down-turned mouths in the middle. Upturned mouths occupy intervening positions on either side of the middle. Eye aperture, with two exceptions, proceeds from narrow at one end, through medial in the middle, to wide at the other end of the dimension. Nor is this the only systematic progression, for eyebrow height progressively decreases from high at the narrow eye aperture end, through medial in the middle, to low at the wide eye aperture end of the scale. Since the third dimension is based on so few triads (only three) it is quite surprising to obtain such systematic relationships. In fact, the only two exceptions in the patterns described

are two faces, F and G, which are not present in any of the three triads upon which the dimension is based. Their relative order is inferred by the attempt to achieve independence for the dimension by not duplicating triad projections on dimensions one and two. Intuitively, the dimension appears independent of the happiness characteristics so prevalent in earlier dimensions. The main effect seems to be a look of increasing determination with greater possibilities for attention as one ascends to the wide eye aperture end of the scale. This is not even intuitively equivalent to the Tension-Sleep dimension which appeared as the third dimension in both Schlosberg's study and our preliminary study. Increasing eye aperture might mediate a tension attribute, but one might expect brow height and mouth curvature would also increase with greater tension. Actually the reverse relationship is found, decreases in these curves corresponding to increases in eye aperture. Eye aperture might also mediate receptiveness, and therefore, with increase in eye aperture, an expectancy or surprise attribute might be anticipated. The dimension yields no such general impression to the experimenter.

The distinctiveness of each type of cue may be assessed intuitively by examining the most consensual triads. The degree of consensus within the group on each triad was recorded (Table 3). For each three stimuli, three orders are possible. Among 13 judges, less than 5 agreements in order on each triad could be expected solely on the basis of chance. Most triads showed considerably greater than chance consensus.

Among those 4 triads showing the greatest degree of consensus (13 agreements out of 13), 3 triads exhibited maximal variation in the mouth curvature cue,of stimuli at either end of the ordered triad, 3 triads

Table 3.

The Degree of Consensus Among Judges on Each Triad Order

CHG	6	AGD	7	BED	9	CAD	9
BFG	10	CDG	5	ABE	9	BFH	10
CFD	10	BFD	10	DHE	11	GFH	6
BDG	10	AHG	6	DHF	9	BHD	13
EHF	6	BCG	10	EBG	7	BAD	8
CAG	10	EBF	8	ADE	6	ACF	9
AFB	10	BEC	7	BCD	7	ACE	12
CFE	6	CFH	11	EHG	8	BHG	11
EFG	13	GDH	13	ACH	10	ADH	7
FAG	10	FCG	6	BAG	6	EDG	6
ECG	10	DCE	6	AEH	9	ADF	5
AHB	8	BCH	5	EAG	5	CEH	6
BFC	7	EBH	8	DFE	9	CHD	13
AFE	13	AFH	12	ACB	11	FDG	6

exhibited maximal variation in the eyebrow height cue, and only 1 triad exhibited maximal variation in the eye aperture cue. Similarly, examination of 6 triads of lesser consensus (11 or 12 agreements) reveals 4 triads varying maximally in the mouth curvature cue, 3 triads varying maximally in the eyebrow-height cue, and 2 triads varying maximally in the eye aperture cue.

These observations suggest that large variations in mouth curvature cues may mediate the high degrees of consensus obtained in individual triad judgments to a greater extent than the other cues. But such indications should not be over-generalized, as maximal variations in each cue are also involved in the least consensual triads (5 agreements).

Such interpretations as are forwarded here are only intuitive and, as such, should be accepted only with the extreme caution accorded the reports of earlier experimenters in this field. A more specific (and empirical) verbal identification of the dimensions was attempted in Study II.

STUDY II

Purpose

Once the basic dimensions have been obtained (the results of Study I), they must be identified. It is most convenient for theoretical purposes if they can be identified with verbal labels or somehow translated into verbal terms. Unfortunately, no fully satisfactory procedure exists to effect such a translation. Previous experiments faced with this problem have employed diverse methods such as:

1. Intuitive inspections based on face-validity criteria. The experimenter would merely inspect the dimension, perhaps to determine what kinds of stimuli occupied key positions or where certain

intuitively meaningful stimuli lay along the dimension, and arrive at a global evaluation of its nature. The resultant identification is often less a resultant of any empirical procedure than of the experimenter's own predilections.

2. Asking the subjects which dimensions they were using. However, few subjects think they are using any dimension or systematic procedure, much less being able to describe verbally one or many emotion states. The concrete nature of their descriptions rarely goes beyond the immediate physiognomic characteristics and then in no systematic fashion. However, the inability to express or describe need not indicate a lack of systematic characteristics among stimuli underlying the response behavior.

3. Comparison with results found using stronger methods. The criticisms of these methods would seem to weaken their utility as a standard of comparison. This is a serious limitation as such methods perhaps reflect more characteristics of verbal, counting, or matching behavior than of the perception of emotion.

In the present study, efforts were made to identify and label the dimensions used by incorporating within the experiment itself variations in instructional set to some of the subjects. It was expected that this procedure would "shrink" the basis of judgments made by a subject instructed to judge on one dimension to that particular dimension alone. It is quite likely the subject may continue to be influenced by intuitive factors not specifically related to that dimension, especially since it is unlikely that the subject ordinarily restricts himself to one dimension in comparing emotions in other individuals. Yet the procedure may

serve to give some information about which dimensions are affected or enhanced. Coupled with identification of relevant facial cues, this should yield a more justifiable identification of dimensions than earlier methods.

It is often suggested that we as humans often approach certain situations with our own hypotheses. One may be prepared for, or expect to see, people or events in a particular way, and evaluate their relative position on this "attribute." Such attributes may rarely carry explicit verbal tags. However, the effect of the instructional set is to induce a certain orientation or set of hypotheses. Once the perceiver is oriented to expect a certain set of alternatives, he is prepared to discriminate all dimensions consonant with this induced orientation, irrespective of their names.

Selection of Attributes

Because not all dimensions are explicitly available or even acceptable to all perceivers, it is necessary to be selective about those dimensions to be employed. Practical considerations further limit the number of dimensions we can study individually to not more than 16, without sacrificing comparability.

Since earlier studies indicated a distinct inability on the part of the perceivers to use bipolar dimensions, only single concepts have the prospect of any unambiguous meaning to each single perceiver.

In the course of these studies many possibilities for attributes have been suggested by the comments of subjects themselves as well as by previous experimenters and theories of emotion. Of all those possibilities suggested, only those attributes meaningful (in the sense that

the subjects did not consider the attributes inapplicable) to a majority of the subjects were retained. This required rejection of words unfamiliar to the subject in the context of judging emotion. Also, concepts of an excessively specific nature had to be rejected since they were not systematic enough to have meaning in comparison of all stimuli. That is, concepts such as "gleefulness" may arise with respect to particular stimuli, yet the subjects may be unable to rate all stimuli on such a restrictive attribute.

Since it would be preferable, for comparison with earlier data, to attain bipolar dimensions, an attempt was made to include for each concept, a hypothetical "logical" opposite in the total set of 16 concepts to be studied. It is preferable if this can be done without merely negating the original concept. This latter procedure would probably first suggest the concept which we would later like to compare. That is, a concept such as "unjoyous" probably only has meaning in reference to the original concept, "joyous." Few negative concepts are in such common use as to allow systematic comparison over a wide range of stimuli.

Subjects

The subjects were 48 (26 male, 22 female) students, selected in the same manner as in Study I, with a comparable age range. Three subjects were randomly assigned to each of the 16 experimental groups.

Procedure

The procedure employed in this study was essentially the same as in Study I with the following modifications.

An additional instruction (Appendix D), to judge on a particular attribute which was specified for each perceiver, appeared on their

instruction sheets. It was deemed inadvisable to attempt complex verbal description or to demonstrate examples to the subjects, since this would unduly complicate the task by adding potentially inconsistent (as well as a priori) elements.

In order to evaluate the subject, rather than the experimenter's ability to formulate and communicate, a single word was employed as a stimulus attribute, with no further explanation. Although some subjects may have been uncertain about using such an attribute, perhaps for the first time in their life, all accepted the conditions of the experiment.

Group triads were then constructed for each of the 16 experimental conditions. Each of the 16 sets of group triads were independently analyzed by the Multidimensional Unfolding Technique. The results were then compared with Study I, allowing a tentative identification of the basic dimensions by comparison with the dimension obtained by subjects instructed to judge on particular attributes.

Results for Study II

Since particular attributes of judgment were specified in this study, the expectation was that perceptual judgments would be somehow channeled by the prior specification of attributes. However, in every case, judging on a particular attribute yielded three dimensions. Three dimensions are also obtained in the analysis of individual as well as group triad data. Therefore, this result is not an artifact of the group summation procedure. There is even a general tendency for more initial triads (those triads which each dimension is chosen to describe) on the third dimension when attributes are specified (Table 4), than in the general, or unspecified case (Study I).

These dimensions (Table 4) can then be individually examined to determine how closely they may yield each general dimension. This can be accomplished by inspection. As a measure of their correspondence, one can determine the number of inversions in order. Thus between each two dimensions which are to be compared, a value for X (inversions in order) can be found as a measure of disagreement.

Table 4.

Dimensions Obtained from Groups Instructed to Judge on Specific Attributes

Aroused	Initial	Final	Relaxed	Initial	Final
1. C A F E B H D G	29	29	1. B C G A D F E H	34	34
2. A G C B D H F E	18	29	2. E B D A F G C H	16	33
3. B E D C F A G H	6	24	3. E H B G A D C F	5	27
Active			**Passive**		
1. C E B H F D A G	30	30	1. C B A F E G H D	37	37
2. E C A G H D B F	17	27	2. A C B D G E F H	10	23
3. B E D C H G A F	6	26	3. E G D A B C H F	6	17
Pleasant			**Unpleasant**		
1. E B H C F D A G	38	38	1. E B F H C D A G	37	37
2. A B C G F E D H	11	22	2. A C G F E D B H	12	24
3. D C E A G F B H	6	21	3. C B H A E F G D	6	17
Approaching			**Avoiding**		
1. B E H C F D A G	42	42	1. C E B F A H D G	33	33
2. C E B F H A G D	12	31	2. A G C F D E B H	18	30
3. A F E C G B D H	2	18	3. B D G E H C A F	4	20

Table 4. Continued

Playful	Initial	Final	Serious	Initial	Final
1. A G D C F H B E	44	44	1. E B F C H D A G	29	29
2. C F A D G E B H	9	19	2. A G C F D E B H	14	27
3. E D F C B A H G	3	20	3. D G F A B E C H	9	24
Artificial			**Genuine**		
1. E B C H F A D G	38	38	1. C A F E G B D H	32	32
2. A D C E H B G F	14	25	2. B E F C G H A D	16	25
3. C B D G F E A H	2	9	3. E D C B A F H G	6	15
Deep			**Superficial**		
1. B E H C F A G D	39	39	1. A G C D F E B H	34	34
2. A C E B H D G F	12	19	2. D G H E A C B F	14	21
3. A H G C D B F E	4	17	3. C D A B H F E G	7	17
Enduring			**Momentary**		
1. B E F C H A D G	37	37	1. E B C F H A D G	40	40
2. A C E B G D F H	16	24	2. A C G D F E B H	12	23
3. A E G C F H D B	3	26	3. C H E A F G B D	4	20

The initial triads in Tables 2.and 4. are those triads which the dimension is chosen to satisfy. The additional triads included in the total number of final triads are those triads which project on previous dimensions.

Kendall's (29) tau coefficient, $\tau = 1 - 2X/\binom{n}{2}$, where n is the number of stimuli in the rank order, may be used to estimate the statistical significance of the similarity between each two dimensions. Since

τ is approximately normally distributed for n = 8, a rough estimate of significance is possible. A one-tailed test of significance can be utilized as the dimensions are recovered as undirected rank orders. The values for X and τ are listed in Tables 2, 5, 6, 7 and 8.

Examination of the evidence for the existence of bipolar attributes (Table 5) reveals that only three pairs of the anticipated "logical" opposites, Pleasant-Unpleasant, Playful-Serious, and Enduring-Momentary, coincide in their principal components in anywhere near the degree necessary to infer bipolar attributes between the two opposite concepts. The evidence for bipolar attributes, in general, is too slight to endure the additional ambiguity which might be introduced by their formulation in this manner.

Examination of the correspondence of each general dimension with each specific attribute dimension (Tables 6, 7 and 8) reveals that these specific dimensions can be arranged in order of decreasing similarity from each general dimension (Table 9).

The first general dimension, D_1, is quite closely approximated by instructions to judge on attributes of Pleasant, Unpleasant or Playful. Furthermore many other specific dimensions approach close correspondence with D_1. Apparently, this primary component permeates almost all judgments even when the subject is specifically instructed to judge on a particular attribute presumably unrelated to Pleasantness. This confirms the expectations that a Pleasantness-Unpleasantness dimension would appear as a primary dimension - as it has in several earlier studies.

Table 5.

Similarity of Dimensions: Specific Attribute Opposites

	Values for X			Values for T		
	1	2	3	1	2	3
Aroused with Relaxed	12	12	14	.14	.14	.00
Active with Passive	9	12	9	.36	.14	.36
Pleasant with Unpleasant	2	5	14	.86	.64	.00
Approaching with Avoiding	7	13	7	.50	.07	.50
Playful with Serious	3	5	13	.79	.64	.07
Artificial with Genuine	13	8	11	.07	.43	.21
Deep with Superficial	7	12	10	.50	.14	.29
Enduring with Momentary	2	6	10	.86	.57	.29

It should be noted that failure to obtain shrinkage to a single dimension or to increase the number of triads satisfied on a primary dimension must be considered as a major limitation in the use of this procedure for identification.

At present, there exists no satisfactory explanation for this failure as very little is known about the "behavior" of complex multidimensional structures. Much further work is necessary to establish the generality of any results from Study II or to specify and locate their source.

Table 6.

Similarity of Dimensions: General with Specific Attributes

D_1 with	Values for X 1	2	3	Values for τ 1	2	3
Aroused	9	6	9	.36	.57	.36
Relaxed	10	12	11	.29	.14	.21
Active	4	14	9	.71	.00	.36
Passive	13	10	12	.07	.29	.14
Pleasant	1	11	11	.93	.21	.21
Unpleasant	1	8	14	.93	.43	.00
Approaching	2	7	13	.86	.50	.07
Avoiding	7	6	12	.50	.57	.14
Playful	1	11	9	.93	.21	.36
Serious	2	6	8	.86	.57	.43
Artificial	3	11	13	.79	.21	.07
Genuine	12	7	10	.14	.50	.29
Deep	4	12	8	.71	.14	.43
Superficial	5	12	13	.64	.14	.07
Enduring	4	12	10	.71	.14	.29
Momentary	4	6	12	.71	.57	.14

Table 7.

Similarity of Dimensions: General with Specific Attributes

D_2 with	Values for X 1	3	2	Values for r 1	3	2
Aroused	8	9	12	.43	.36	.14
Relaxed	10	11	10	.29	.21	.29
Active	10	11	12	.29	.21	.14
Passive	5	12	10	.64	.14	.29
Pleasant	9	11	7	.36	.21	.50
Unpleasant	9	6	12	.36	.57	.14
Approaching	8	9	7	.43	.36	.50
Avoiding	7	12	14	.50	.14	.00
Playful	11	13	13	.21	.07	.07
Serious	8	10	12	.43	.29	.14
Artificial	7	13	13	.50	.07	.07
Genuine	10	10	7	.29	.29	.50
Deep	6	13	8	.57	.07	.43
Superficial	13	13	6	.07	.07	.57
Enduring	6	10	8	.57	.29	.43
Momentary	6	12	14	.57	.14	.00

Table 8.

Similarity of Dimensions: General with Specific Attributes

D_3 with	Values for X 1	2	3	Values for r 1	2	3
Aroused	10	6	11	.29	.57	.21
Relaxed	6	14	11	.57	.00	.21
Active	14	8	13	.00	.43	.07
Passive	7	4	12	.50	.71	.14
Pleasant	11	5	9	.21	.64	.36
Unpleasant	11	6	10	.21	.57	.29
Approaching	14	11	9	.00	.21	.36
Avoiding	11	8	14	.21	.43	.00
Playful	9	9	13	.36	.36	.07
Serious	12	8	13	.14	.43	.07
Artificial	13	11	9	.07	.21	.36
Genuine	6	13	12	.57	.07	.14
Deep	14	8	11	.00	.43	.21
Superficial	7	12	11	.50	.14	.21
Enduring	14	4	10	.00	.71	.29
Momentary	14	7	12	.00	.50	.14

All values of $r \geq .57$ ($X \leq 6$) are significant at the .05 level of confidence.

Table 9.

Primary Attributes Ordered as to Similarity to Each General Dimension

D_1	X	D_2	X	D_3	X
Pleasant	1	Passive	5	Relaxed	6
Unpleasant	1	Deep	6	Genuine	6
Playful	1	Enduring	6	Passive	7
Approaching	2	Momentary	6	Superficial	7
Serious	2	Avoiding	7	Playful	9
Artificial	3	Artificial	7	Aroused	10
Active	4	Aroused	8	Pleasant	11
Deep	4	Approaching	8	Unpleasant	11
Enduring	4	Serious	8	Avoiding	11
Momentary	4	Pleasant	9	Serious	12
Superficial	5	Unpleasant	9	Artificial	13
Avoiding	7	Relaxed	10	Active	14
Aroused	9	Active	10	Approaching	14
Relaxed	10	Genuine	10	Deep	14
Genuine	12	Playful	11	Enduring	14
Passive	13	Superficial	13	Momentary	14

In the case of the second general dimension, D_2, an examination of the principal component, or first dimension, in each specific attribute group, reveals a less clearcut correspondence. The closest attributes one can infer are Passive, Deep, Enduring, Momentary, Avoiding and Artificial. Thus D_2 appears different from D_1 in spite of any intuitive similarity in distribution of facial cues (Study I). From past studies, Approaching or Avoiding might be expected to be the most descriptive attributes for D_2. The actual evidence suggests that other attributes not in the list might have been more appropriate. The only bipolar combination in the list which shows closer correspondence in the principal component than Approaching-Avoiding is Enduring-Momentary.

The third dimension has, as its closest attributes, Relaxed, Genuine, Superficial and Passive. Here again, no attribute is close enough to be a useful description of the dimension, and words external to the list might have achieved greater success. It should be noted, however, that two of the first four attributes are semantically akin to a Tension-Sleep scale, particularly the latter or Sleep end of the scale. Since the existence of the Sleep end of the dimension was in question, the present study may indicate that such a scale is descriptive of facial expressions of emotion. If one were to select a bipolar attribute, the closest correspondence would seem to be Aroused-Relaxed. A more fortunate word choice than Aroused might have yielded a better descriptive fit.

STUDY III

Purpose

Direct comparisons can be made with dimensions inferred from actual facial expression data. Hypothetically, one might expect more dimensions

to be utilized as the complexity of the stimuli is increased by the use of actual human faces.

Subjects

The subjects were 8 (4 male, 4 female) students selected from the same population as in the previous two studies, with a comparable age range.

Procedure

The procedure employed was to read to the subjects the same basic introductory statement (Appendix B), as in Study I. A different instruction sheet (Appendix E) was given them, however, and any questions were answered.

The stimuli used were the eight Frois-Wittman pictures employed in our preliminary study. These eight stimuli were projected on the screen simultaneously, arranged in the same linear order as in each of the dimensions derived in that preliminary experiment. Since three dimensions were determined, three presentations of eight stimuli each were necessitated. The subjects were allowed as much time as necessary to make their judgments and the pictures were continuously visible during this period. Although the task was complex in nature, each period was approximately 10 minutes in duration.

Each subject was provided with a set of eight cards. Each card depicted one of the eight schematic faces in Studies I and II. The nature of the task was to examine the dimension or order scale of human faces projected on the screen, decide what attribute of emotion is involved and then to order the schematic faces along the same scale.

Methods of Analysis

These order scales were collected for each individual. These were converted into three order scales for the group by assigning integers from 1 to 8 to each stimulus corresponding to its order in each individual scale. These numbers could then be summed over the group for each stimulus on each scale. The order of the sums can then be taken as the order of the corresponding stimuli, and the three group scales, for schematic faces can be constructed and compared with the three group scales for human faces.

Results for Study III

To determine if these general dimensions are the same as those used in judgments on human facial expressions of emotion, a direct comparison was effected. The task was to arrange the schematic faces in the same orders as represented by the attributes of human emotion as discovered in a preliminary study. The orders of schematic faces obtained by this process were then compared with the general dimensions. The correspondence, Table 10, was generally quite close, except for the second dimension, which seems to be rather ambiguous in all the studies.

Table 10.

Scales Obtained by Matching the Dimensions of Human (Male) Emotion Found in Preliminary Study, and Degree of Correspondence

		X Values			Order of Correspondence
		D_1	D_2	D_3	
M_1	A G D C B F E H	4	11	6	D_1, D_3, D_2
M_2	H B E F C G D A	5	11	9	D_1, D_3, D_2
M_3	H E D G B F A C	10	6	5	D_3, D_2, D_1

RELIABILITY

The results obtained in Studies II and III raise the question of the reliability and representativeness of the general group dimensions obtained in Study I. That is, to what extent are the dimensions which may be obtained by individual analyses (the Multidimensional Unfolding Technique applied to the sets of judgments obtained from each subject) consistent with each other? To what extent are the group dimensions representative of the dimensions used by each individual? This type of information could enable one to estimate reliability by the degree of inter-subject consistency.

The dimensions obtained by individual analyses are listed in Table 11, along with the general group dimension to which each corresponds most closely.

Table 11.

Individual Dimensions

Individual	I_1	D	I_2	D	I_3	D
1	A C F G B E H D	3	G D C A F H B E	1	C B H A D F G E	2
2	A G D C F H B E	1	G A C B E F H D	3	B A C E H D F G	2
3	D G A H F C E B	2	A C B E F H D G	2	C A G B E D H F	3
4	A G D C F H B E	1	B E C F A H G D	2	C A B G E D H F	3
5	A G C D F H B E	1	B E F A C H D G	2	C A G F E B D H	3
6	C F E B H A G H	2	A G D C F H E B	1	C A G B E D F H	3
7	G D A F C H B E	1	A B E C F H G D	2	C A G B E D F H	3
8	G A D C F H B E	1	A B C E F G H D	2		
9	G A D H F C B E	1	C A B G E D F H	3	B A E H C G F D	2
10	G A D F C H B E	1	B A C E F H G D	2	C A G E B D F H	3
11	G D A F C H E B	1	A C G B E D H F	3	B E A F C H G D	2
12	G D A F H C B E	1	B E A F C H D G	2	C A G B D E F H	3
13	G A D C F H B E	1	B A F C E G H D	2	C A G E B D F H	3

From Table 11, it can be seen that eleven of the thirteen dimensions extracted first in the individual analyses agree significantly with the general group dimension D_1. Eight of the thirteen dimensions extracted second correspond to D_2. Eight of the twelve dimensions extracted third correspond to D_3. Twelve of the individual analyses yielded three dimensions, while in a single case only two dimensions were required. It should be noted that although most of the subjects used dimensions similar to each of the group dimensions, their order of extraction was often not the same as for the group as a whole. This may indicate that the respective importance of each dimension in each individual's perceptual set may vary some across individuals.

Measures of correspondence (in terms of τ) between each individual dimension and the group dimensions are presented in Table 12. In each case, a significant τ between <u>some</u> dimension in each of the individual analyses and the dimension D_1 of the group analysis was found. Eleven of these measures were significant beyond the 1 percent level, and the other two were significant beyond the 5 per cent level. Similarly, for D_2, seven of the thirteen measures were significant beyond the 1 per cent level, and four were significant beyond the 5 per cent level. For D_3, nine of the twelve measures were significant beyond the 1 per cent level, and three were significant beyond the 5 percent level. For each individual, a dimension corresponding most closely to D_1 may be selected and the τ measure of correspondence noted. These measures can be summed and averaged across individuals, yielding an average measure of correspondence, along with an estimate of statistical significance. These are shown

in Table 13. Each group dimension demonstrates a highly significant degree of correspondence with the set of corresponding (maximally similar) individual dimensions.

Table 12.

Correspondence of General Group Dimensions with Individual Dimensions

	Correspondence								
	D_1 with			D_2 with			D_3 with		
Individual	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
1	.07	.86	.00	.43	.43	.14	.57	.29	.14
2	.93	.36	.14	.21	.36	.71	.36	.64	.29
3	.64	.14	.21	.71	.71	.21	.14	.43	.93
4	.93	.36	.21	.14	.79	.36	.36	.14	.86
5	.79	.57	.29	.14	.71	.14	.36	.07	.64
6	.36	.93	.29	.36	.29	.29	.07	.21	1.00
7	.86	.21	.29	.43	.93	.29	.14	.36	1.00
8	1.00	.07	--	.29	.79	--	.14	.50	--
9	.79	.21	.36	.50	.36	.79	.07	.93	.21
10	.93	.21	.21	.36	.93	.14	.21	.36	.93
11	.79	.21	.43	.50	.29	.86	.07	.79	.14
12	.79	.43	.36	.50	.79	.21	.07	.07	.93
13	1.00	.14	.07	.29	.71	.21	.29	.29	.93

All values of $r \geq .57$ are significant at the .05 level of confidence.

All values of $r \geq .74$ are significant at the .01 level of confidence.

Furthermore, for purposes of comparison, average measures of correspondence can be calculated and examined for significance between each group dimension and individual dimensions not exhibiting maximal similarity. That is, of the thirteen individual dimensions extracted first only eleven were maximally similar to D_1. It is possible to examine the two individual's first dimensions not exhibiting maximal similarity to D_1, average their measures of correspondence with D_1 and test their significance. This general procedure was followed for all group dimensions, with the results presented in Table 14. If these correspondences were significant, it would seem that the relationship obtained between general and individual dimensions might be attributable to their relative order of extraction as much as to any unique identity of each dimension.

With one possible exception, significance estimates clustered about the .50 level (a high possibility of obtaining such a degree of correspondence on a basis of chance alone). This would suggest that the small agreements found among group and individual dimensions which are not maximally agreeing could have been obtained on the basis of chance alone. No measure in Table 14 attained a .05 level of significance.

Table 13.

Average Correspondence of General Group Dimensions with Individual Dimensions

	D_1	D_2	D_3
Average (mean) $\bar{r}$	.856	.693	.813
$\chi^2 = \frac{\Sigma(r-0)^2}{\sigma_r^2}$	107.7	86.5	109.1
Degrees of Freedom	13	13	12
Significance Level	$p << .001$	$p << .001$	$p << .001$

Table 14.

Average Correspondence of General Group Dimensions with Individual Dimensions Non-maximal in Similarity

	D_1			D_2			D_3		
	I_1	I_2	I_3	I_1	I_2	I_3	I_1	I_2	I_3
$\bar{r}$	.215	.264	.256	.344	.340	.216	.187	.270	.192
χ^2	1.60	11.92	9.43	19.02	7.15	5.75	7.14	11.13	1.93
D.F.	2	11	12	12	5	9	12	10	4
p	.50	.50	.70	.10 (>.05)	.30	.80	.90	.50	.80

Even though the group dimensions correlate highly on the average with the component individual dimensions, it is possible nevertheless that the representative group dimensions may be misleading as to their consistency of usage between subjects. The average amount of inter-subject consistency may be calculated by the formula (24)

$$\bar{r} = \sum_i \frac{(a_i)^2}{\frac{1}{2}\, n(n-1)k(k-1)} - \frac{1}{(k-1)}$$

where a_i = the difference between the numbers of individuals agreeing and disagreeing on the order of pair i.
n = the number of stimuli ranked.
k = the number of individuals.

If this amount of consistency on each dimension in the order of extraction were too low, it would be hazardous to assume that all the individual structures with different orders of dimensional extractions could be represented adequately by a single group structure with a single order of extraction of dimensions. These average consistency estimates (Table 15) are reasonably high, except for the second dimension extracted. These values are considerably improved if only those dimensions agreeing maximally with a given group dimension are considered.

Table 15

Average Interagreement between Individuals' Dimensions.

In individual order of extraction.		Among those agreeing most highly with group dimensions.	
I	.612	I	.733
II	.183	II	.632
III	.450	III	.644

The average or group dimension is apparently a "good estimate" since, even though the average level of inter-subject consistency is high, the average correspondence with the group dimension is higher (Table 13) than the average agreement among subjects. Thus the group dimensions seem to be best representative of the individual dimensions. In addition, this evidence leads to the conclusion that there is reliability, in the sense that each of the subjects seemed to be basing their judgments, in general, upon the dimensions extracted from the group triads. Nothing is guaranteed about the order of extraction of these dimensions, however reliable.

In resumé, three dimensions have been located and can be fairly well identified by the distribution of mediating physical cues. A precise verbal characterization is more elusive, however. A more fortuitous choice of words or stimuli might have "rotated" the specific attribute dimensions into a better fit with the general dimensions. Until such a rotation can be effected, however, any definite interpretation must remain primarily intuitive. The general dimension D_1 may perhaps best be described as primarily an evaluation of pleasantness. D_3 can be described as an attribute of relaxation or other factors related to the immediate intensity of emotion. D_2 seems to be an attribute primarily related to temporal factors. These factors may refer to duration or anticipatory characteristics of emotion.

Chapter VI

SUMMARY AND IMPLICATIONS

In summary, several studies of judging emotion from facial expression were examined and found to have crucial limitations. A greater control of stimulus variations, both within a single face and within a series of faces, was indicated. Response possibilities were often too constricted and insufficiently analyzed.

The present study utilized symmetric schematic faces as stimuli. Three cues - mouth curvature, eye aperture and eyebrow height - were varied. Each cue had three possible values. On the basis of preliminary studies, 8 of the 27 possible schematic faces were selected as stimuli.

In Study I, 13 introductory psychology students judged each of the 56 possible triads by the Method of Similarities. These judgments were analyzed by a multidimensional technique of "unfolding" analogous to non-metric factor analysis, resulting in 3 dimensions.

Study II attempted the identification of the 3 general dimensions by repeating the first study with 48 different subjects and specifying particular attributes of judgment. Comparison of the resultant particular dimensions with the 3 general dimensions identified the first general dimension as Pleasant-Unpleasant. The other general dimensions were more difficult to identify. The third general dimension was closest to Aroused-Relaxed, and the second dimension was closest to Enduring-

Momentary or Approaching-Avoiding. The degree of correspondence was attenuated by the insufficient evidence for the existence of bipolar attributes.

In Study III, scales composed of Frois-Wittman human poses of emotion, ordered along dimensions, were presented to 8 subjects. These dimensions of human faces suggested 3 dimensions of schematic faces which corresponded fairly close with the general dimensions obtained in Study I. The only exception to this close correspondence was the second or Attention-Rejection dimension.

The same 3 dimensions seem to recur in roughly the same form throughout the present studies of perception of emotion. Perhaps these 3 dimensions could serve as a step toward an operationally defined descriptive basis for judgment of emotion.

Implications for Future Study

This attempt to establish and operationally define a descriptive basis for emotion should not be taken to imply an elementalist point of view. Past builders of classification systems have often assumed that their categories described, in some sense, "elements" of emotion out of which more complex emotions were physiologically compounded. We have already seen the fallacy of trying to foist a system of concepts at one level onto a different level, resulting in hypothetical and inadequate description.

In contrast, the present dimensions are merely inferences from perceptual judgment response data, obtained from an external observer, of a possible expression of emotion. Thus they could not be expected to correspond to clear-cut somatic correlates in a person feeling the emotion,

since there is no a priori reason to expect any combination of behavior, even at the physiological level, to be "unidimensional" in nature.

Each dimension must be expected, then, to be the resultant of many psychological and physiological events and processes. A psychological dimension may or may not prove fruitful in physiological investigations by covarying with manipulation of an experimental variable. But its value is not contingent on this demonstration of a hypothetical covariation. It suffices to determine a model for the reliable description and classification of emotions at an observational level, which can serve as one criterion for experimental manipulation and enable one to achieve additional discrimination in the analysis both of complex emotional response, and response of a perceiver to the cues of emotion in another.

Certainly no claim is being advanced that "all" emotion or even all perception of it can be reduced to these three dimensions. Entirely different dimensions may be found in, say, introspective experience. But it is up to those who wish to communicate about such dimensions to find and specify them, and to demonstrate their reliability and accessibility to observation or manipulation. The fact the 19th century introspectionists such as Wundt (55) proposed and found evidence for roughly the same three dimensions (Lust-Unlust, Erregung-Beruhigung and Spannung-Losung) in introspective experience of emotion suggests that there is little difference. Arrival at these same three dimensions from data in other microareas, however, increases our confidence in their wide range of applicability. But, it remains an individual decision whether one can utilize this system in any particular area. In such a case, equipped with such

a system, one can look for new dimensions of behavior and experience, and introduce new order into observations, both inside and outside the laboratory. This can be done without sacrificing meaning and comparability of concepts, or relational aspects of emotion.

It is interesting to note the development of the concept of feeling. In early Wundtian psychology, feeling was conceived of primarily as a feeling-tone of sensation, and a non-essential attribute of sensation at that. Later, analysis of feeling itself was attempted and it gradually began to acquire an independent existence as an element of consciousness.

Kulpe (32) objected to the conception of feeling as an attribute because it was not an essential attribute; that is, it had unique qualities of its own. He also claimed that an attribute should in itself be irreducible as an aspect (no attribute should have attributes of its own). This new conception was unfortunate in several respects. The study of feeling and emotion lost any objective references to sensation and foundered in a morass of conflicting, unresolvable introspective evidence. Its conception as an irreducible aspect discouraged research on its multidimensional structure and led to fruitless controversy. Its conception as an element rather than an attribute encouraged premature objectification with a resultant overly specific emphasis on empirical research.

With the lapse into subjective introspection, the Pleasantness-Unpleasantness dimensions apparently became so salient, or so emphasized in reference to the self,that any further dimensions became almost indetectible.

The present study endeavors to restore some objective properties to the study of the cues of subjective affective phenomena. Areas of irrational content need not always be approached with irrational methods. This study

attempts to extend investigation of referred emotion from Wundtian analyses of sensation to an analysis of perception on a behavioral basis. Further extensions to more complex material, involving greater conceptual involvement are conceivable.

Such an extension might eventually explore and extract enough "variance," in reference to objective factors, to reveal characteristically individual cognitive structures of either a permanent or temporary nature. Even in its present form the technique could detect how variations in a subject's mood can effect his perception of emotion in others. This could be generalized to a clinical test. Once adequate concepts of emotions have been explicitly formulated we will possess normative criteria, with which we can compare perception of emotional behavior in undeveloped or pathological states to determine more exactly these specific kinds of deficit in ability to perceive emotional differences which may be involved in certain types of functional and organic pathology. It has often been suggested that people who are mentally ill are deficient in this respect, and it may well be that the disturbance in ability to perceive emotional difference is responsible for the changed concept of the self in its relations with the external world which so often characterizes states of mental illness.

Certainly more variation in stimuli should be undertaken. Starting from the present limited study of schematic facial expression a larger sample of possible stimulus variation should be investigated. With the progressive inclusion of more human facial characteristics one may be able to build up a more extended picture of perceived emotion as it is related to human events. An investigation which samples emotion cues

over an extended time period would be especially enlightening on temporal cues in the inference of emotion, which were difficult to locate in the present study. Three dimensions were obtained when this same type of study was performed on human faces (first preliminary experiment) as well as in the present study of schematic faces where only three cues are varied. There is no a priori reason to expect that the same three dimensions would be preserved in studies employing stimuli intermediate in complexity.

In Study III, two of the dimensions seemed to preserve identity. However, the third, or D_2, intuitively seemed to the experimenter to involve a quality of "expectancy" which may be radically different in nature in the present schematic faces than in human faces. That is, the particular aggregate and pattern of cues employed may communicate a quality which is in itself varying, at least over the range from schematic faces to human faces. It is possible that human faces involve a greater degree of observer-supplied context or interaction between stimuli and observers. Perhaps as progressively more cues are employed, a transition point can be found where schematic faces can yield a D_2 similar to the second dimension (M_2) yielded by human faces.

Similarly, as progressivelyfewer cues, or cues that are more "homogeneous", are employed, further alterations and perhaps the loss of a dimension might be expected. As stimuli involving more cues than are transmitted by photographs are employed, more dimensions might be expected.

The present studies suffer from the limited number of cues which were varied in a systematic manner. The connection between the utilization of three variable cues and the emergence of three dimensions of judgment is not clear. Further studies involving the cues varied here imbedded in a larger set of variable cues would help to resolve this issue.

Just as the use of standardized schematic faces and the Multidimensional Unfolding Technique were exploratory innovations in the study of perception of emotion, the method of identification used in Study II was primarily exploratory. The prior specification of particular attributes could not be expected to "rotate" dimensions to a concentrated, simple and unidimensional structure unless several or most of the stimuli constituted exemplifications of extremes of only that particular dimension corresponding to the verbal concept specified. Further stimulus inputs, such as situational or verbal constraints, can affect the dimensions of perception in various ways, not entirely understood. Seemingly innocuous variations in reference can affect tremendous changes in the conception and structure of referred emotion.

These variations in stimulus input can operate as constraints, or change entirely the nature of the response. In this study, an effort was made to reduce the restrictions imposed on the perceiver to a primitive level. Judgments of similarity (or more properly, most alike and least alike) were considered to involve a minimum of suggested or imposed attributes or restrictions. In actual practice, considerable variation is sometimes found between judgments of similarity and judgments of dissimilarity, which are not a priori symmetrically inverse predicates. The relation of the two judgments probably varies somewhat with the stage of concept formation in the subject. Unfortunately, among the suggestions of this study is one that conceptual judgments of emotion receive little aid from verbal systems and often have not yet achieved this property of systematic conceptual reversibility. This is probably the same sort of difficulty as is faced in relating discrimination learning to generalization in traditional investigations of learning processes.

However, the Method of Similarities is one of the most justifiable methods of investigation on a phenomenal basis. The chief requirement in its employment is a more precise specification of the response task,

i.e., whether to judge similarity or dissimilarity and in what sequence, and to communicate this instruction to the experimental subject. Then some analytic tool such as Hays' extension of Coombs' Unfolding Technique, as used here, may be applied with a minimum of restrictive assumptions.

It should be noted that while this system of analysis is less restrictive than factor analysis in defining dimensions (since the dimensions are defined by their end-points which are stimuli in the set), it also employs definite assumptions as to the nature of the judgmental process and the dimensions which may be obtained. These dimensions, while not units in a customary sense, form related wholes along which different phenomena may be ordered, thus requiring transitive, antisymmetric relations to be valid between elements in the domain. Although these assumptions seem modest enough, still weaker systems may be utilized, perhaps uncovering other dimensions. Or stronger systems employing symmetric or metric assumptions may be used. Which system we actually employ depends on the assumptions we are prepared to make and what properties we wish to require of the dimensions.

This recourse to such weak techniques may seem to be too severe for present needs. But the present conceptual state in perception of emotion (and, indeed, in the larger area of social perception) demonstrates a profound need for revision of some basic assumptions. One must face problems of identity, similarity, order and stable consensual descriptions before we can even attempt measurement in its exact sense. Perhaps eventually quantification along these dimensions in an area besides that of facial expression data will be demonstrated to be meaningful, as, for example, in perceived gestures, or introspected emotional states.

It is quite likely there are more dimensions to emotion than are discoverable through analysis of facial expressions. Since this sample of behavior has been severely restricted so as to facilitate comparability, the inferences from this data should not be construed as exhaustive but rather as constituting a minimal set. The determination of these three dimensions in such a restricted sample suggests that we are no longer justified in regarding more complex phenomena as unidimensional without investigating the relative independence of factors.

At this point it is interesting to note that Osgood (37), in determining by more conventional factor analytic methods the dimensions in semantic space along which descriptive adjectives may vary, has found three dimensions, very similar to those of this study, of evaluation, potency and activity. Since his procedure was largely designed to reflect connotative meanings prevalent in our language, it is surprising to achieve similar dimensions and in the same order of appearance as with more denotative facial expression data.

Such results present us with an interesting dilemma, the solution of which might clarify much of the confusion reigning in the study of language about the nature of meaning, and in emotion about the relation of descriptions. Do we use these particular constructs in understanding emotion because the influence of our language on thought is so profound (even on supposedly nonverbal behavior)? Or, is the influence of emotion on our lives so strong that its dimensions color the way we can use language or associate to it? Only further steps into this uncharted area may bring us in sight of an answer.

APPENDIX A

Faces*	Scales** 1	2	3	4	5	6	7	8	9
ULHM	4.0	4.4	4.6	4.7	3.6	4.7	4.2	4.0	amused
ULMM	3.2	3.3	2.9	3.3	2.8	3.4	3.2	2.9	amazed
ULMW	4.1	4.4	4.6	4.3	3.5	4.6	3.8	3.7	
ULLM	3.9	4.0	4.0	4.0	3.3	4.1	4.3	4.0	happy
ULLW	3.1	3.4							
ULHW	4.0	4.2	4.5	4.4	3.5	4.5	3.7	3.8	happy
USHM	3.7	4.3	4.2	4.1	3.3	4.2	3.8	4.0	pleased
USHM	2.9	3.2	3.2	3.4	2.9	3.6	3.0	2.7	surprise
USMM	3.6	3.5	3.9	3.8	3.2	3.8	3.9	3.8	
USMW	3.7	4.0	3.7	3.7	3.3	3.8	3.4	3.2	cheerful
USMW	3.2	2.9	3.4	3.8	3.4	3.1	2.9	2.6	surprised
USLM	3.3	3.7	3.4	3.7	3.1	3.5	3.6	3.3	satisfied
USLM	2.8	2.7	3.3	3.3	3.2	3.7	3.9	3.6	happy
USMM	4.0	4.1	3.6	3.7	3.1	3.7	4.0	3.6	happy
USLW	2.5	1.8							angry
SLHM	3.1	2.6	2.7	2.7	1.9	2.6	2.3	2.2	placid
SLHW	3.6	3.2	2.8	3.3	2.7	2.9	2.7	2.9	questioning
SLMM	3.2	2.1	2.3	2.7	2.5	2.5	2.4	2.3	
SLMW	3.0	2.3							indifferent
SLLW	3.0	2.4							indifferent
SLLM	2.6	1.7	2.8	2.9	3.5	2.5	3.1	2.7	angry
SSHW	3.6	3.1	2.9	3.0	2.6	2.9	2.8	2.7	complacent
SSHM	3.2	2.3	2.5	2.7	2.6	2.8	2.9	2.1	sad
SSHM	2.7	1.5	2.3	2.3	2.5	2.2	2.0	1.8	worried
SSMM									
SSMW	2.7	2.1							fright
SSLM	3.2	2.2	2.3	2.3	2.4	2.6	2.7	2.5	complacent
SSLW	2.0	1.5	1.7	2.6	3.5	2.0	2.6	2.5	furious
SSMM	3.4	2.5	2.6	2.6	2.2	2.9	2.8	2.7	indifferent
DLHM	3.1	1.5	1.2	1.8	3.4	1.3	1.9	2.8	sad
DLHW	3.2	1.6							sad, scared
DLMM	3.1	1.6	1.1	1.6	3.6	1.7	1.9	3.1	sad
DLMW	3.1	1.3	1.1	1.8	3.5	1.4	1.9	2.8	sad
DLLM	3.2	1.3	1.2	1.7	3.8	1.3	1.9	3.1	sad
DLLW	3.1	1.5							sad
DSHM	2.6	1.6	1.7	1.8	2.7	1.8	2.0	2.3	sad
DSHW	3.2	2.4							sad
DSMW	3.0	1.8	1.8	2.4	3.4	1.8	2.6	2.8	sad
DSMM	3.0	1.6	1.5	1.6	2.9	1.6	2.1	2.5	sad
DSLW	3.0	1.8	1.7	2.0	3.5	1.8	2.8	3.1	sad
DSLM	2.8	1.5	1.3	2.0	3.6	1.3	2.8	3.0	angry

*In the abbreviations at the left, the first letter signifies mouth curvature - Up, Straight or Down. The second letter signifies mouth length - Long or Short. The third letter signifies eyebrow height - High, Medium or Low. The fourth letter signifies eye aperture - Wide or Medium.

**1. Frequency
2. Preference
3. Pleasant-Unpleasant
4. Attention-Rejection
5. Tension-Sleep
6. Approach-Avoid
7. Organized-Disorganized
8. Aware-Unaware
9. Typical Verbal Labels

APPENDIX B

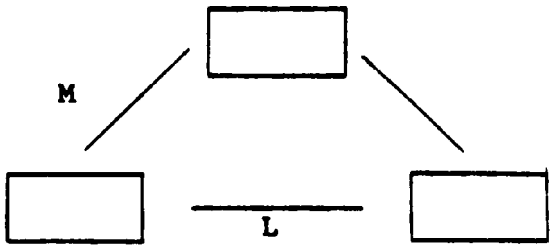
We have been conducting a series of experiments to determine how well people can perceive emotion from facial expression. This is not at all as easy a problem as it sounds. There are many assumptions and difficulties involved and we've been forced to consider many variations in procedure.

In particular, tonight we'd like to see how well you can judge emotion from a series of schematic facial expressions which we shall present on the screen.

Now read the instruction sheet and if you're not clear about anything or have any questions, raise your hand.

APPENDIX C

It is a well known fact that facial expressions can be good indicators of the feelings of an individual. We often infer how an individual feels in a particular situation by his facial expressions. We want to study how well you can judge these feelings from the following pictures. These pictures will be presented three at a time. On the following pages make your judgments. For example, in the following diagram



the box at the top will stand for the picture at the top of the screen, the box at the left will stand for the picture at the left of the screen, and the box at the right will stand for the picture at the right of the screen.

On the lines interconnecting the boxes you are to make your judgments. Put an M on the line between the two pictures that are <u>most</u> alike with respect to the feelings of the individual. Put an L on the line between the two pictures that are <u>least</u> alike with respect to the feelings of the individual.

You will have 20 seconds to make your judgments on each set of pictures. While performing this task it is requested that you maintain silence.

APPENDIX D

Different emotions can vary in several respects - for example, they may differ in their qualities or attributes. There are many attributes or dimensions on which different emotions may be compared. We would like you to compare different emotions on one particular attribute - that is, make your judgments of most alike and least alike with respect to the attribute below.

IMPORTANT - be sure that you keep this attribute in mind while judging. Make all your judgments on this basis alone.

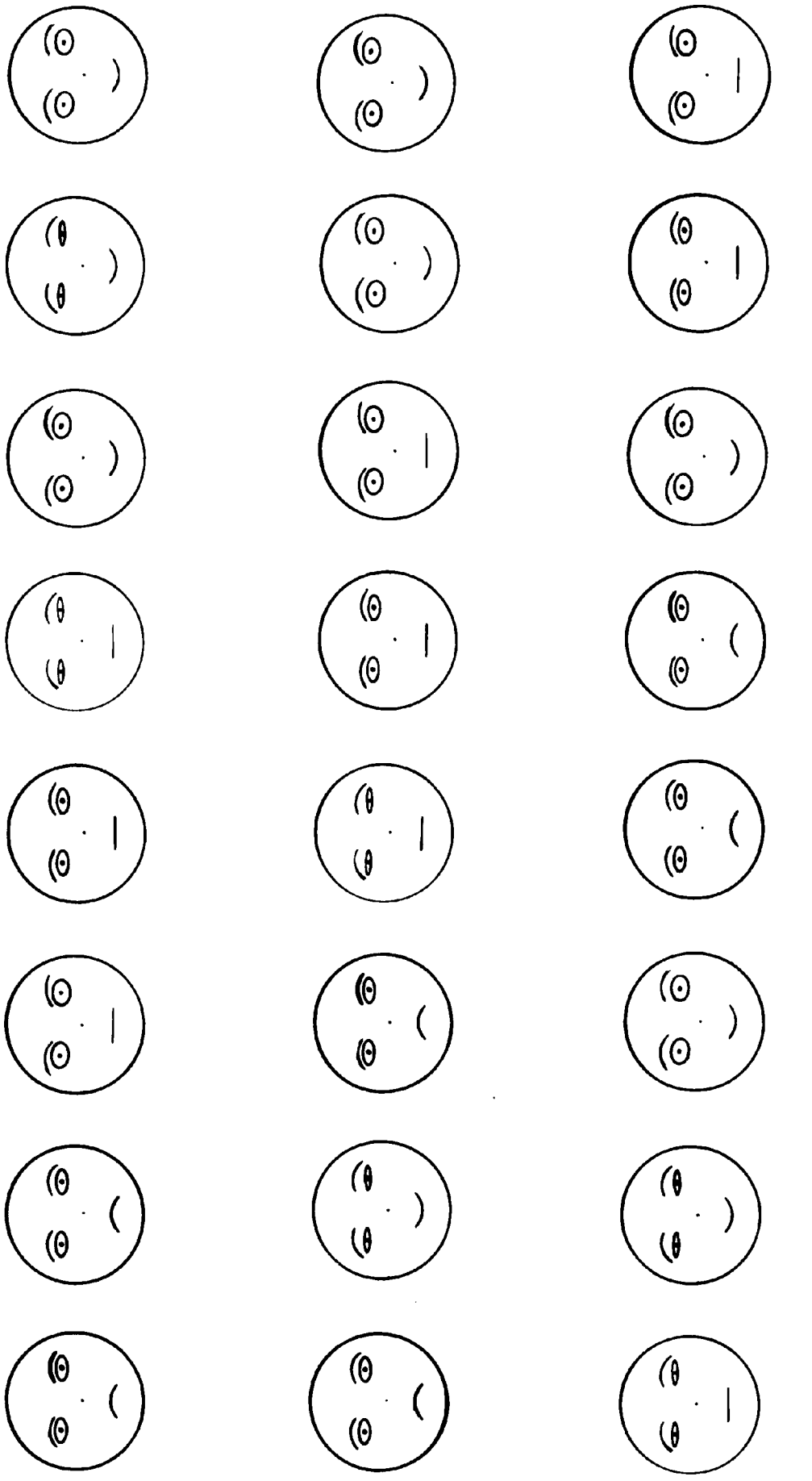
Attribute: Judge with respect to how the emotion seems to be.

APPENDIX E

There are many ways in which different emotions may be compared. Not only are there many qualities or attributes they may have in common, but an emotion may also have more, or less, of a certain quality than another emotion. Such an ordering of emotions on the basis of possessing more or less of a certain attribute is often known as an order scale.

Examine each of the order scales of facial expressions of emotion as they are projected on the screen. Try to see how they vary and what attribute of emotion determines their order on the scale. Then, on the dotted line beneath each scale, order the individual faces along the same scale.

For example, given the following scale and a set of individual faces, designated by letters, if the faces seem most appropriate in the order H G F E D C B A, you would indicate this on the dotted line below the scale.



D_1 D_2 D_3

BIBLIOGRAPHY

1. Allport, F.H. Social Psychology. Boston: Houghton-Miflin, 1924.

2. Allport, G.W. Personality: A Psychological Interpretation. New York: Henry Holt, 1937.

3. Allport, G.W. and Odbert, H.S. Trait-names: A psycho-lexical study. Psychol. Monogr., 1936, 47, No. 1. (Whole No. 211).

4. Boring, E.G. and Titchener, E.B. A model for the demonstration of facial expression. Amer. J. Psychol., 1923, 34, 471-485.

5. Bruner, J.S. and Tagiuri, R. The perception of people. Chptr. 17 in Lindzey, G. (Ed.) Handbook of Social Psychology. Cambridge, Mass.: Addison-Wesley, 1954.

6. Brunswik, E. Perception and the Representative Design of Psychological Experiments. Berkeley: Univ. of California, 1956.

7. Brunswik, E. and Reiter, L. Eindrucks-Charaktere schematisierter Gesichter. Z. f. Psychol., 1937, 142, 67-134.

8. Bühler, C. The social behavior of children. In Murchison, C. (Ed.) A Handbook of Child Psychology (Rev. Ed.). Worcester, Mass.: Clark Univ., 1933.

9. Busby, D.E. The interpretation of facial expressions. Amer. J. Psychol., 1924, 35, 602-604.

10. Classe, A. The whistled language of La Gomera. Scientific American, 1957, 196, 111-120.

11. Coleman, J.C. Facial expressions of emotion. Psychol. Monogr., 1949, 63, No. 1 (Whole No. 296).

12. Coombs, C.H. A method for the study of interstimulus similarity. Psychometrika, 1954, 19, 183-194.

13. Coombs, C.H., Raiffa, H., and Thrall, R.M. Some views on mathematical models and measurement theory. Chptr. 2 in Thrall, R.M., Coombs, C.H. and Davis, R.L. (Ed.). New York: Wiley, 1954.

14. Darwin, C. The Expression of the Emotions in Man and Animals. London: Murray, 1872.

15. Delafresnaye, J.F. (Ed.) Brain Mechanisms and Consciousness. Oxford: Blackwell, 1954.

16. Duffy, E. The concept of energy mobilization. Psychol. Rev., 1951, 58, 30-40.

17. Feleky, A.M. The expression of the emotions. Psychol. Rev., 1914, 21, 33-41.

18. Feleky, A.M. Feelings and Emotions. New York: Pioneer, 1924.

19. Fernberger, S.W. False suggestion and the Piderit model. Amer. J. Psychol., 1928, 40, 562-568.

20. Frijda, N.H. The understanding of facial expressions of emotion. Acta Psychol., 1953, 9, 294-356.

21. Frois-Wittman, J. The judgment of facial expression. J. exp. Psychol., 1930, 13, 113-151.

22. Gates, G.S. An experimental study of the growth of social perception. J. educ. Psychol., 1923, 14, 449-462.

23. Guilford, J.P. An experiment in learning to read facial expression. J. abnorm. soc. Psychol., 1929, 24, 191-202.

24. Hays, W.L. Personal communication.

25. Hays, W.L. A proposal for the non-metric analysis of multidimensional similarities data. Dittoed unpublished manuscript.

26. Hulin, W.S. and Katz, D. The Frois-Wittman pictures of facial expression. J. exp. Psychol., 1935, 18, 482-498.

27. Jenness, A. The recognition of facial expressions of emotion. Psychol. Bull., 1932, 29, 324-350.

28. Kanner, L. Judging emotions from facial expressions. Psychol. Monogr., 1931, 41, No. 3 (Whole No. 186).

29. Kendall, M.G. Rank Correlation Methods. London: Griffin, 1948.

30. Klineberg, O. Emotional expression in Chinese literature. J. abnorm. soc. Psychol., 1938, 33, 517-520.

31. Kremenak, M. Der Eindruckswert der Augengegend auf Grund schematischer Darstellungen. Univ. of Vienna Doctoral Dissertation, 1950.

32. Külpe, O. Grundriss der Psychologie. Leipzig: Engelmann, 1893.

33. Landis, C. Studies of emotional reactions: II. general behavior and facial expression. J. comp. Psychol., 1924, 4, 447-509.

34. Langfeld, H.S. The judgment of emotions from facial expressions. J. abnorm. soc. Psychol., 1918, 13, 172-184.

35. Lindsley, D.B. Emotion. Chptr. 14 in Stevens, S.S. (Ed.), Handbook of Experimental Psychology. New York: Wiley, 1951.

36. Munn, N.L. The effect of knowledge of the situation upon judgment of emotion from facial expression. J. abnorm. soc. Psychol., 1940, 35, 324-338.

37. Osgood, C.E., Suci, G.J., and Tannenbaum, F.H. The Measurement of Meaning. Urbana: Univ. of Illinois, 1957.

38. Piderit, T. Mimik und Physiognomik. Detmold: Meyer (H. Denecke), 1886.

39. Ruckmick, C.A. A preliminary study of the emotions. Psychol. Monogr., 1921, 30, No. 3 (Whole No. 136).

40. Rudolph, H. Der Ausdruck der Gemutsbewegungen des Menschen. Dresden: Kuehtmann, 1903.

41. Ruesch, J., and Klees, W. Nonverbal Communication. Berkeley: Univ. of Calif., 1956.

42. Schlosberg, H. A scale for judgment of facial expressions. J. exp. Psychol., 1941, 29, 497-510.

43. Schlosberg, H. The description of facial expressions in terms of two dimensions. J. exp. Psychol., 1952, 44, 229-237.

44. Schlosberg, H. Three dimensions of emotion. Psychol. Rev., 1954, 61, 81-88.

45. Secord, P.F. in Tagiuri, R. and Petrullo, L. Person Perception and Interpersonal Behavior. Stanford Univ., 1958.

46. Secord, P.F., Dukes, W.F., and Bevan, W. Personalities in faces: I. an experiment in social perceiving. Genet. Psychol. Monogr., 1954, 49, 231-279.

47. Sherman, M. The differentiation of emotional responses in infants: I. judgments of emotional responses from motion picture views and from actual observation. J. comp. Psychol., 1927, 7, 265-284.

48. Spitz, R.A. and Wolf, K.M. The smiling response: A contribution to the ontogenesis of social relations. Genet. Psychol. Monogr., 1946, 34, 57-125.

49. Stern, W. General Psychology from the Personalistic Standpoint. New York: Macmillan, 1938.

50. Stevens, S.S. Mathematics, measurement, and psychophysics. Chptr. 1 in Stevens, S.S. (Ed.), Handbook of Experimental Psychology. New York: Wiley, 1951.

51. Winkler, M. Der Ausdruckswert der Mundgegend auf Grund schematischer Darstellungen. Univ. of Vienna Doctoral Dissertation, 1951.

52. Woodworth, R.S. How emotions are identified and classified. Chptr. 18 in Reymert, M.L. (Ed.), Feelings and Emotion: The Wittenberg Symposium. Worcester, Mass.: Clark Univ., 1928.

53. Woodworth, R.S. Experimental Psychology. New York: Henry Holt, 1938.

54. Woodworth, R.S., and Schlosberg, H. Experimental Psychology (Rev. Ed.). New York: Henry Holt, 1954.

55. Wundt, W. Grundriss der Psychologie. Leipzig: Engelmann, 1896.