XIII

Communication through nonverbal behavior: A source of information about an interpersonal relationship

Paul Ekman

Re-Printed from:
Affect, Cognition and Personality
Edited by Silvan S. Tomkins & Carroll E. Izard
Springer Publishing Company, Inc., New York

XIII

Communication through nonverbal behavior: A source of information about an interpersonal relationship

Paul Ekman

INTRODUCTION

The interlacing of fingers, twist of a foot, slump of a shoulder, slant of a hip, curl of the lip, furrow of the brow, direction of gaze, and tilt of the head are instances of what we have called nonverbal behavior. All can occur simultaneously or separately, with or without speech, during an interaction or when an individual is alone, spontaneously or by contrivance. The concern of our study is with the communicative functions of this domain of behavior, when actions are spontaneously emitted during a verbal exchange with an immediate, emotionally-toned interpersonal relationship.

Poets and politicians, psychotherapists and playwrights, dancers and anthropologists have all provided examples of and testimonials to the rich variety of information which can be carried by interactive nonverbal behavior. The claims vary from the modest hypothesis that nonverbal cues provide qualifications about how a verbal message should be interpreted, to notions that this mode of communication escapes conscious censoring and thus reveals the "true," primitive, or repressed side of personality.

Without subscribing to the belief that body movement and facial expression are a royal road to the unconscious, or even to affect, our

This research was supported by a postdoctoral research fellowship, #MF 6092, and by a grant, #MH 07587, both from the National Institute of Mental Health, United States Public Health Service. Parts of these studies were reported at the Western Psychological Association Convention in 1962, 1963, and 1964, and at the American Psychological Association Convention, 1964.

The author is indebted to Dr. Robert Berryman, who initially encouraged this line of investigation, and to Mr. Sanford Autumn for his many criticisms and suggestions about the research design, data analysis, and interpretation. The author is grateful to Patricia W. Garlan for her editorial assistance with both the style and the content; to Joanna Bressler, who helped in the design, collection of data, and data analysis in Experiment VI and in the experiment with hospitalized depressed patients. and to Marvin Hoffman, who helped in the design, collection of data, and data analysis in Experiment VII.

decision to study *interactive* nonverbal behavior is based on the assumption that, at least in part, nonverbal behavior expresses the quality and changes of a relationship. As Bateson (1962) has suggested, this mode of expression is especially sensitive to the nuances and intricacies of how two people are getting along, despite the possibility that they, and we as observers, customarily pay little attention to this channel. Our emphasis has been on exploring what is communicated about a relationship when attention is fixed on this source alone.

Nonverbal behavior has been the subject of experimental research in psychology since the 1920's—in the early days under the rubric of expressive behavior, more recently within the framework of person-perception, emotion, and the interview. There are a number of reviews of this literature. For a recent and quite extensive review, see Klein (1963). Other recent reviews are those of Allport (1961), Bruner & Tagiuri (1954), Brengelmann (1963), and Davitz (1964). Remarkably, few of the experiments have studied *interactive* nonverbal behavior, and their relevance is limited by the setting and methods employed to elicit the nonverbal behavior.

The most popular technique has been to have actors, professional or amateur, pose certain emotions or reactions to imagined events. When individuals are so specifically instructed to send a wide range of emotions, and so strictly limited to the use of the nonverbal channel, it would appear to us that the behavior shown will be extreme or at least atypical of expressions occurring during an interpersonal transaction. Any demonstration of communication of information through nonverbal behavior in such a situation is not surprising, and is of doubtful relevance to the question of whether less artifically induced nonverbal cues communicate accurate information. The few studies which have sampled spontaneous behavior provoked the subject with novel, extreme, or bizarre stimuli, which again do not provide a picture of a person's usual nonverbal repertoire.

In either case, sampling nonverbal behavior occurring outside the context of an on-going relationship, devoid of any continuing verbal exchange, removes an important and perhaps critical constraint. Rarely do people express anger or joy unrestrictedly; usually the nature of the interpersonal relationship imposes constraints on the manner and extent of expression. A further problem with these kinds of experiments is that usually the relationship—between an experimenter giving instructions or producing novel stimuli, and a subject either acting or reacting—is blurred by the focus on task-relevant behavior.

Before discussing more recent studies which have examined spontaneous interactive nonverbal behavior, let us distinguish between two approaches, the *indicative* and the *communicative*. A nonverbal act, such as a foot tap, can be shown to be an indicator of an internal or external event, or to be a communicator of a specific item or class of information; but these are two quite different questions which entail different methods of study.

In indication, the concern is not with what a group of receivers may

observe but with the relationship the experimenter is able to establish between a nonverbal act and some other class of events. Thus, the frequency of foot taps might be related to a verbal theme, or the administration of a drug, or the stress in an interview; and foot taps would then be an indicator of this other variable. Indication studies require a method of describing or recording specific classes of nonverbal activity. (See the notational system of Birdwhistell (1952); criticism of this system by Ekman (1957); more recent approaches by Buehler & Richmond (1963), Dierssen, Lorenc, & Spitalerl (1961), Jones et al. (1955, 1958, 1961), Sainesbury (1954). Studies of nonverbal indication examine only the sender within the communication system and tell us nothing directly about whether a receiver can decode any systematic information from a nonverbal indicator. For example, while foot taps might have been found to occur with greatest frequency when the sender is wistfully recalling the pleasures of early childhood, there is no reason to think that this indicator has communicative value, that the untutored receiver would infer this verbal theme, or related affect, from observing a sender tap his foot.

Communication through a nonverbal act is established only by determining whether receivers agree in their observations or in their inferences about what the act portends. Typically, a communicative study entails presenting segments of nonverbal behavior to groups of receivers who act as judges, and measuring the consistency or accuracy of their responses. Communication can be inaccurate as well as accurate; for example, if we believe that people smile only when happy, then this behavior has communicative value to receivers even if careful indicative experiments were to demonstrate that senders frequently smile under stress. Accuracy or inaccuracy of communication can be examined only if, of course, there is some independent criterion relevant to the sender's experience or his intentional attempt to communicate. Without such a criterion, communicative studies can still investigate whether a nonverbal act or series of acts provides information that is consistently interpreted (whether rightly or wrongly) or is ambiguous.

Failure to find that a nonverbal act has communicative value does not necessarily preclude the possibility that the act does have indicative value. The act may be a very strong indicator, but not familiar to the group of receivers being sampled, or perhaps not normally interpreted unless a receiver is trained to look for it. A communicative design does not provide an exhaustive sifting of what may be communicated through nonverbal behavior; it may be only a first approximation of what is potentially available.

Determination that a nonverbal act has indicative and/or communicative value does not assume that the sender intended to communicate. What the sender intends, and what the experimenter discovers through an indicative approach, or a receiver infers in a communicative study, can be completely unrelated.

Similarly, what the sender actually experiences, intentionally or non-

intentionally, is not only almost impossible to determine, but is in no way necessarily equivalent to the indicative or communicative value of his nonverbal act. Resting the finger on the nose might have the communicative value to receivers, "Things smell bad." Or it might have the indicative value of occurring with greatest frequency at those moments when expert ratings of the interview typescript suggest maximal relaxation. And yet it may be that a given subject puts his finger to the side of his nose precisely at those moments when he is finished with a particular line of discourse, as an anticipatory signal that he will now ask a question.

There have been a number of recent studies of spontaneous interactive nonverbal behavior using an indicative approach. Examining clinical interviews with a patient, Dittmann (1962) found patterns of body movement indicative of the patient's mood as assessed independently by experts. More recently, Dittmann (1963) found consistencies in the pattern of body movements for each of a number of different normal subjects studied over time. Utilizing an EMG to measure body movement, Sainesbury (1955) found that the amount of movement was indicative of stress within a structured interview, of disturbance, verbal themes, and specific affects, as determined by psychiatric ratings. Exline (1963) found that amount of eye contact during interviews is indicative of reactions to embarrassing themes, or competitiveness, and bears a relationship to the sex of the subject.

Findings in comunicative studies of interviews have been more contradictory. Let us first consider two studies of interactive nonverbal behavior outside the interview situation. Investigations of experimenter bias by Rosenthal (1963a, 1963b) have found that much of the transmission of bias between experimenter and subject is carried through nonverbal cues, without awareness. At present, he is examining films of these experimenter-subject interactions to specify the particular nonverbal cues involved. Maccoby, Jecker, et al. (1964), and Jecker, Maccoby, et al. (1964) have studied students' nonverbal behavior during classroom lectures, have found nonverbal indicators of whether a student understands the lecture; and in communicative studies using teachers as receivers they have through training increased teacher accuracy in interpreting the students' behavior.

Turning now to communicative studies of nonverbal behavior within an interview, two studies varying widely in results may be cited. Mahl (1959), acting as the sole receiver, inferred with startling accuracy information about the emotional state, diagnostic classification and psychodynamic features of groups of patients being interviewed. His judgments were based solely upon observation of the nonverbal channel and then were validated by comparison with the verbal content of the interview and the case history. In marked contrast to this report is a study by Giedt (1955, 1958) which found that nonverbal cues were of little value in making clinical postdictions about four patients. Comparing judgments of an interview made on the basis of typescripts, tape-recordings, silent film, and sound film, he found that the most accurate inferences were

made on the basis of the verbal cues as compared with nonverbal cues, with little improvement from a sound-film presentation of stimuli as compared to a typescript presentation of only the verbal content. The surprising results of this research raise a number of questions about the nature of the accuracy criteria, the suitability of the judgment task to the types of information which can be communicated nonverbally, the possibility that the patients photographed were not nonverbal senders, that the situation did not elicit sufficiently varied nonverbal behavior, and that the judges may not have been skilled as nonverbal receivers. Nevertheless, Giedt's study did serve to focus attention on a number of methodological considerations which directly influenced our choice of research problem and method of investigation.

Problem

In summary, then, most of the early research on expressive behavior and many of the recent studies of person-perception and affect treating nonverbal behavior were considered largely irrelevant to our interest in studying the communicative functions of nonverbal behavior spontaneously emitted during a verbal exchange in which two people have an emotional involvement in their relationship. A distinction was drawn between two kinds of approach, the indicative and the communicative. By 1959, when the present research was planned, there had been increasing evidence that spontaneous interactive nonverbal behavior is *indicative of* a number of different types of information about a person, but there was a clear contradiction between Giedt's results and Mahl's as to whether such behavior during interviews *communicates* accurate information. Our series of experiments was addressed to this discrepancy, and focused upon whether nonverbal behavior communicates accurate information about the quality of an interpersonal relationship.

The decision to study the ways in which nonverbal behavior reflects changes in a relationship was based upon the assumption that this is an important function of this communication channel, and also upon considerations as to the reasons for Giedt's failure to produce positive results about nonverbal behavior as a source of information. The two major considerations involved the sampling of the sender's behavior and the type of judgment asked of the receiver.

To provide a reasonable test of what is communicated by spontaneous nonverbal behavior, the sample should include diverse but representative pictures of the sender's nonverbal repertoire. If the sender is shown in only one mood, dealing with only one type of task, in a relationship characterized by a single relationship quality, then the receiver is deprived of the opportunity to compare and contrast possibly different nonverbal cues which might be available if mood, task, or relationship quality were sufficiently varied. The need then to sample different experiences of the sender, yet to stay within the artificial constraint of one photographing

session of a single interview, led us to decide upon a standard stress interview procedure, which will be discussed shortly.

The receiver's task, the determination he is asked to make from the sender's behavior, is also crucial. Obviously, certain types of judgment task will result in one hundred percent accuracy (e.g., to determine the sender's sex), while other types of judgment task will elicit no consistent results whatsoever (e.g., to determine the sender's earliest memory). The judgment task assigned to the receiver, then, reflects the experimenter's beliefs about what may potentially be found to be communicated through nonverbal behavior. The indicative studies of Exline, Dittmann, and Sainesbury, and the writings of many theorists, in particular Sullivan (1953), suggest that what the patient is saying verbally and the nature of the relationship he is experiencing are related to his nonverbal activity. In another separate series of experiments (Ekman, 1964a), we found that there is congruence between verbal information and nonverbal information communicated at the same moment in time. Now our purpose was to move beyond the specifics of the verbal exchange, and concentrate upon whether a change in the over-all relationship between two interactants would be accurately communicated through nonverbal behavior.

Assuming that we could demonstrate that nonverbal behavior communicates accurate information about the quality of an interpersonal relationship, we were interested in investigating three further more specific questions: (1) How much of the nonverbal behavior shown during an interview actually carries consistent information; how much of the behavior is ambiguous to the receiver? Are the moments when the nonverbal activity carries information relatively rare, or is this a more continuous reiterated phenomenon? In other terms, what is the signal-to-noise ratio? (2) What specific classes of information are communicated by nonverbal behavior? How does a receiver decode a nonverbal act into some determination about the relationship between two people? Does the communication of information about the relationship depend upon some awareness of the circumstances of the interaction, or is such communication possible without most situational or contextual aids? (3) If individual senders differ in their nonverbal communicativeness, can the responses of a group of receivers serve as the basis for distinguishing between senders in the clarity of their nonverbal communicativeness?

The interview relationship is one in which the behavior of at least the interviewee is fairly spontaneous, there is an explicit communicative process, verbal behavior is the primary method of discourse, and the constraints operate against obtaining rich or extreme nonverbal records. In order to sample nonverbal behavior adequately within different interpersonal relationships, a standard interview procedure was designed in which the interviewer's attitude and behavior could be manipulated so as to permit two grossly different interactions within a short time span.

A second decision concerned the method of recording the nonverbal behavior. There are disadvantages to each of the three options: still photographs, motion pictures, and no records. Still photographs present the

behavior to receivers in an artificial fashion, and are usually employed in designs in which only a few isolated slices of behavior are to be judged. Motion pictures are expensive to acquire, and often overwhelm the investigator with a mass of stimuli as complex as the original behavior, from which he must either sample only selected pieces for scrutiny or obtain judgments based on such a conglomerate mass of stimuli that it is impossible to specify how many cues, and of what type, form the basis of the receiver's judgment. If no records are made, then the receivers must be present when the behavior occurs, replication is difficult at best, and specification of what cues and how many cues served as the basis for the judgment is usually not possible.

Our decision was to record time-sample still photographs, a compromise dictated by practical considerations to limit cost but still provide a sufficient number of photographs to represent much of the behavior which occurred, and to permit us to specify how many cues and of what type communicate information. The price of this compromise, however, was that we were limited to investigating positions rather than movements, with consequent artificiality in the presentation of the sender's behavior to the receivers. Further, in studying the question of how much of the behavior shown during an interview communicates consistent or ambiguous information, we are limited by our stimulus unit of a still photograph to learning only how many of these photographic samples communicate consistent or ambiguous information.

A third decision was to photograph the entire face and body of the sender, rather than limit ourselves to recording only one portion of the sender's nonverbal activity. This decision was based on our notion that different parts of the body transmit somewhat different types of information, and therefore the receiver should have the total nonverbal message at a moment in time available.[1]

A fourth decision concerned the judgment task. The task selected was simply that judges distinguish, from photographs, the two standardized interaction patterns; in other words, judges were to determine whether a picture had been taken when the interviewer acted toward the subject in one specified way, or in the other. This task had a deceptively easy accuracy criterion which will be the subject of a much later discussion.

The fifth, and last, general decision concerned the nature of the judge population. College freshmen were chosen, not only because of their availability, but also because of our interest in determining what is communicated to a naive, untrained receiver. Results from pilot studies were in agreement with the findings reported in the literature, that experts,

[1] The usual stimulus unit was a single photograph showing all of the observable nonverbal behavior of the sender at one moment in time. By obtaining judgments for many such units, we could then determine how many of these units communicated consistent information. In other research (Ekman, 1964b) we have used a smaller stimulus unit, subdividing the single photograph into head or facial cues, and body cues, to compare the information communicated by each. But this will not be reported in this chapter. The logical next step, further to subdivide and analyze the different cues within the face and within the body, is in progress.

at least in the mental health professions, are no more and perhaps less accurate judges than college freshmen.

A number of experiments were conducted. In each of the experiments different groups of judges responded to nonverbal stimuli gathered from a series of standardized interviews. The specifics of the interview procedure, the interview participants, methods of recording the nonverbal stimuli, the judgment task and instructions to the judges are described in detail and followed by description of the individual experiments. Summaries of this material are on pages 405 and 414. After the second summary, the general discussion explores the three questions outlined (how much information is communicated, what kind of information is communicated, and how can individual differences in sending clarity be measured), and closes with a speculative discussion of the range of information which further research may show is communiated through nonverbal behavior.

EXPERIMENTS

GENERAL METHOD

Interview Procedure

A standard rather than a clinical interview was used, to better elicit contrasting affective reactions. Both the style and content of the interviewer's behavior toward the interviewee, the subject, were programmed. After an introductory affectively neutral period of ten minutes the interviewer induced a period of stress by attacking and criticizing the subject. After ten minutes the interviewer initiated a catharsis phase of ten minutes by explaining the experiment to the subject, and praising and joking with him. (The interview structure is described more fully in the Instructions to the Judges, presented below in Experiment I.)

Interview Participants

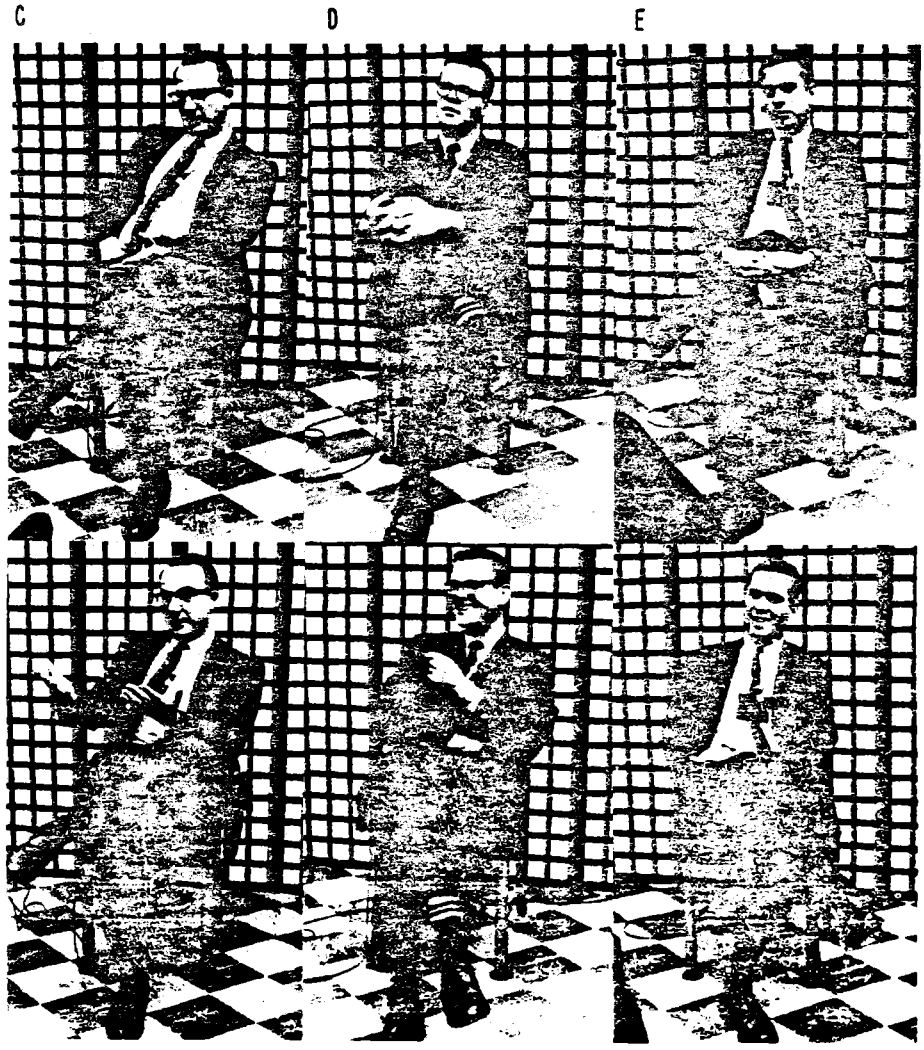
Five interviews were conducted, in which three different interviewers and five different subjects participated.[2] Two interviewers were staff research psychologists at a Veterans Administration hospital; each interviewed one psychology graduate student (Interviews A and B). The third interviewer, a senior staff research psychiatrist at a different neuropsychiatric hospital, interviewed three different psychiatric residents (Interviews C, D, and E). The interviewer in each case was in a position of some authority over the subject. The subject knew that he was participating in an experiment and that the interview would be tape-recorded and observed, but did not know that he would be photographed. The interviewer was acquainted with the general purpose of the research, and had rehearsed the interview procedure. (The five interviews will hereafter be referred to as A, B, C, D, and E.)

[2] The author is grateful for the participation of Drs. John Boswell, Enoch Calloway, Barry Decker, Irving Moelis, Joseph Rubinstein, David Saunders, Lee Wannerman, Robert Weiss.

The choice of these interview subjects, trainees in psychology and psychiatry, was dictated by ethical and methodological considerations. The stress experience is perhaps not totally unrewarding to persons who are trainees in the professional use of interview procedures; learning about this technique may partially compensate for their suffering as subjects. It was possible to utilize the knowledge of the clinical staff to select trainees

Figure 1. Examples of photographs. Top: stress, bottom: catharsis.

able to tolerate a stress experience of this nature. The chief disadvantage of this choice was that conceivably trainees in psychology or psychiatry are less active or expressive nonverbally than most other populations, and that the constraints of being interviewed by a senior staff person might further limit the range of their nonverbal behavior. Viewed methodologically, however, such limitations were held to be advantageous, in



that they provided a more rigorous framework for obtaining positive results.

Recording Method

Photographs were taken through a one-way vision screen with a 35mm still camera. In interviews A and B the photographs showed a profile view of both interviewer and subject. In C, D, and E the photographs showed a full-face view of the subject only, with the camera aimed approximately over the interviewer's shoulder. Pictures were taken every thirty seconds during A, every fifteen seconds during B, and every five seconds during C, D, and E. (Interviews C, D, and E were recorded about a year later than A and B, when it had become evident that greater frequency of sampling was desirable, and more elaborate equipment had become available.) Examples of these photographs are shown in Figure 1, with the interviewer in A and B cut off.

Judgment Task

Judges were asked to guess whether a photograph had been taken during the stress or catharsis portion of the interview. Although it is quite possible that judges may have considered the affective experience of the subject in making their decision, it is important to note that judges were not asked to judge the subject's affect, but instead to decide during which phase of the interview a picture had been taken. The two phases were defined in the instructions primarily in terms of the nature of the relationship between interviewer and subject.

Since pilot studies had shown that judges misconstrued the word "catharsis" in the instructions, the two phases of the interview were designated as the "stress" and "final" phases. Judges saw photographs which had been randomly selected without prior inspection.

EXPERIMENTS I-IV

Experiment I

Problem. The purpose of this experiment was to determine whether nonverbal behavior communicates accurate information about two interview conditions described for the judge in terms of two different interactions between interviewer and subject. Hypothesis: Untrained judges can accurately decide whether a photograph was taken during the stress or catharsis phase of an interview.

The most stringent test of this hypothesis demanded that judgments be free of (a) the influence of practice, (b) base-line information about the subject's typical behavior, (c) comparative information about the range of behavior shown during stress and catharsis, and (d) information about sequence of positions over time. Therefore, the design restricted each judge to seeing only *one* photograph. In this form the judgment task is quite different from the usual way in which people judge others, since

many of the circumstances which aid interpretation of nonverbal behavior have been eliminated. The virtue of the design is that it allows exploration of what can be learned from a minimum of nonverbal cues.

Method. A total of 48 photographs from A and B were presented to judges, half from the stress phase of the two interviews and half from the catharsis phase. Fourteen photographs were from A: all seven of the photographs taken during the catharsis phase, and seven photographs randomly selected from the 19 taken during the stress phase. Thirty-four photographs were from B: 17 randomly selected from the 40 taken during the stress phase, and 17 randomly selected from the 31 taken during the catharsis phase. The 48 photographs, showing a profile view of both interviewer and subject, were enlarged into 4 x 5" prints. A photograph was randomly assigned to each judge as he entered his classroom. Judges were given the following instructions to read:

> This is a study of the way in which people are able to interpret or understand gestures, body movement, and facial expression. You will be shown a photograph which was taken during a short standardized interview. The interviewer, or Examiner, was a staff psychologist at a hospital, and the person interviewed, the Subject, was a student in training to become a psychologist. The Subject was told that he was participating in a research project on interviewing techniques and that he would be observed through a one-way vision screen and tape recorded.
>
> Both of the interviews were standardized in that the Examiner followed a prearranged schedule of behavior designed to evoke two different emotional reactions in the Subject. After the first few minutes of getting acquainted, the Examiner became hostile and challenging, disagreeing with everything the Subject said, and continually interrupting him. With one Subject the Examiner questioned his motives for being interested in research; with the other Subject the Examiner concentrated more on the Subject's "poor" preparation for his coming examinations, and the reasons why the Subject claimed to be interested in the physiological areas of psychology. After ten minutes of the *STRESS*ful period, the Examiner explained what he had been doing, that it had been part of the research to try to provoke the Subject and study his reactions to stress. In this *FINAL* period, the Examiner attempted to reassure the Subject, and generally bring about a release of tension.
>
> In actuality, this standardized interview plan did not work out perfectly in either of the two interviews. There was some stress for the Subject throughout the interview. The *FINAL* phase was not completely successful in producing relief from the stress, since the Subject knew that the experiment was still continuing. Nevertheless there were some important differences between the two phases of the interview. The *STRESS* phase did have more overt expression of hostility and tension; and in the *FINAL* phase the Subject did experience some relief, at least in knowing that the worst was over.
>
> You will be given one picture face down. Please do not look at it until you are asked to do so. When you are instructed, turn the picture over and study it. You will be allowed 20 seconds in which to make your judgment as to whether you think the picture was taken during the *STRESS*ful or the *FINAL* period of the interview. At the end of the 20 seconds you will be

asked to write your judgment, either the word *STRESS* or the word *FINAL* at the bottom of this page. Next to your judgment please write the number of the picture which you will find on the reverse side in the top corners.

Subjects. Forty-eight college freshmen served as judges.

Results. A total of 48 independent judgments was obtained, one for each photograph. It was assumed that if only chance factors had been operative half of the judgments would be correct. A binomial test (Siegel, 1956, p. 36) was used to evaluate the significance of the number of accurate judgments. Table 1 shows that the total number of judges who made an accurate choice across both interviews (.65) was significantly greater than would be expected by chance. When the accuracy on each interview was tested separately, the number of judges correct on A (.64) was not significant, while the number of judges correct on B (.65) was significant. (The failure to achieve significance on the judgments of A can be attributed to the small number of judges, since the proportion of correct judgments was almost identical to that obtained for B.)

Table 1 Number of Js who made an accurate judgment

	Experiment I 48 Independent judgments Interviewer and S	Experiment II 48 Independent judgments S alone
Total trials (A and B) (48 photos)	31*	24
Interview A trials (14 photos)	9	8
Interview B trials (34 photos)	22**	16

*$p < .05$
**$p < .01$

Discussion. These results supported the hypothesis that judges can accurately decide whether a photograph was taken during the stress or catharsis phase of an interview. Having shown both interviewer and subject in the pictures, however, we could not be sure of the extent to which accurate judgment was based on the subject's behavior, the interviewer's behavior, or inference on the part of the judges about the interaction between interviewer and subject. Experiment II was designed to address this question.

Experiment II

Problem. The purpose of this experiment was to replicate Experiment I, while limiting judges to the nonverbal behavior of only the subject. Hypothesis: Untrained judges can determine whether a photograph was taken during the stress or catharsis phase of an interview even when limited to seeing the nonverbal behavior of the subject only.

Method. The procedure was identical to that of the first experiment; the same photographs were used, except that the part of the picture showing the interviewer was covered with a cardboard mount to eliminate the interviewer's cues.

Subjects. A new group of 48 college freshmen served as judges.

Results. A total of 48 independent judgments was obtained, one for each photograph. It can be seen in Table 1 that the hypothesis was not supported, since only .50 of the judges of A and B combined made a correct choice. The difference in accuracy between Experiments I and II was evaluated by a χ^2 test. While the accuracy on A was not significantly different when only the subject was shown (.58, as compared to .64 in Experiment I), the accuracy on B was less when only the subject was shown (.47, as compared to .65 in Experiment I; $p<.05$).

Discussion. It would appear from the results of Experiment II that when the interviewer's behavior is not shown, judges can no longer make accurate decisions, particularly for B. Earlier it was noted that the design employed in these experiments, allowing only one judgment of one photograph to each judge, deprived the judges of the benefits of practice, familiarity with base-line or comparative information, and sequence cues. Thus handicapped, a moderate number of judges still made a correct decision in the first experiment, although in the second experiment the further elimination of interviewer cues led to chance results. The next two experiments studied how judges' accuracy would be affected when some of these handicaps were removed.

Experiment III

Problem. The purpose of this experiment was to determine whether judges could accurately identify stress and catharsis photographs when each judge made a decision for each of a number of pictures. The procedure allowed judges the opportunity to benefit from practice, and to acquire impressions about the subject's nonverbal repertoire, specifically about the behavior shown in stress and catharsis. Since the photographs were shown in a random order, with a decision required for each picture, judges were still deprived, as they had been in the earlier experiments, of any information based on the sequence of nonverbal behavior shown in the interviews. Hypothesis: Untrained judges can accurately determine, when seeing a series of randomly selected and ordered photographs showing both interviewer and subject, whether pictures were taken during the stress or catharsis phase of an interview.

Method. The same task was employed although the procedure was modified for presentation of a series of photographs rather than one. The instructions to judges were slightly modified (in regard to the method of recording decisions). The judges as a group then saw in a random order the 48 photographs used in Experiments I and II. The pictures had been made into 35mm slides, and each was projected on a screen for twenty seconds. As in Experiment I, the pictures showed both interviewer

and subject. Two versions of the answer sheet controlled for a left-right response-set bias; the word "stress" appeared at the left side of the page on one answer sheet, and at the right side on the other.

Subjects. A new group of 33 college freshmen served as judges.

Results. The number of accurate choices for each judge was tabulated. If the photographs had not systematically provided any information, accuracy scores would be expected to be symmetrically distributed at about the midpoint of the range of possible scores. A contrary distribution of accuracy scores would indicate systematic response by the independent judges to the randomly presented pictures. The statistical hypothesis was tested by applying Wilcoxon's matched-pairs signed-ranks test (Siegel, 1956, p. 75) to the differences between the accuracy score obtained for each judge and the expected midpoint; a one-tailed test was employed. (This method of testing the significance of accuracy results was employed in all of the subsequent experiments unless otherwise noted.)

The accuracy achieved on all trials and on A and B trials separately

Table 2 Results of Experiments III, IV, and V

Pictures correctly identified	Experiment III (33 Judges) Interviewer and S		Experiment IV (35 Judges) S Alone		Experiment V (16 Judges) S Alone	
	Score	Proportion	Score	Proportion	Score	Proportion
Total trials A & B (48 photos)						
Median	29.3***	.61	27.0***	.56	24.0	.50
1st quartile	27.1	.56	24.7	.51	21.7	.45
3rd quartile	32.2	.67	28.4	.59	26.2	.55
Interview A trials (14 photos)						
Median	10.1***	.72	9.7***	.69	7.5	.54
1st quartile	8.7	.62	8.2	.59	5.3	.38
3rd quartile	11.0	.79	10.7	.76	8.4	.60
Interview B trials (34 photos)						
Median	20.1***	.59	17.3	.51	16.5	.48
1st quartile	18.9	.55	16.4	.48	14.2	.42
3rd quartile	21.4	.63	18.4	.54	19.0	.56

*** $p < .001$

was significantly higher than would be expected by chance. The distribution of scores was skewed towards accuracy, particularly for interview A photographs. Table 2 shows the results for Experiment III and for the next two experiments. The median and the first and third quartiles are shown to present information about the nature of the distribution of accuracy scores. Proportions are given in addition to the actual scores, in order to facilitate comparisons between interviews A and B, where different numbers of photographs had been judged.

Discussion. Although direct comparison of the results of Experiments I and III is confused by the difference in design between requiring a single judgment and repeated judgments from each judge, descriptively it can be noted that in both of these experiments in which interviewer and subject were shown significant accuracy was achieved. Moreover, the level of accuracy over all of the photographs from the two interviews was remarkably similar in Experiment I (.65 correct) and Experiment III (.61 correct) despite the fact that the multiple judgment results of Experiment III reflect practice, familiarity with the range of behavior, etc. This similarity in results raises a number of questions about the nature of the judgment task, and how the accuracy of judgment may or may not be influenced by increased exposure of judges to nonverbal stimuli. These questions will be discussed after the first nine experiments have been presented.

The next experiment studied whether accuracy was possible if the interviewer is not shown in the photographs, but with each judge making a decision for each of a number of photographs.

Experiment IV

Problem. Experiment II had suggested that judges could not accurately decide whether a photograph was taken in the stress or catharsis phase when the pictures showed only the behavior of the subject. Experiment IV, like Experiment III, gave the judge opportunity for practice and for acquiring comparative information about the behavior shown in stress and catharsis; as in Experiment II, judges saw only the subject's behavior. Hypothesis: Untrained judges can accurately determine from a series of randomly selected and ordered photographs whether pictures were taken during the stress or catharsis phase of an interview even when limited to seeing the nonverbal behavior of the subject only.

Method. The procedure was identical to that of Experiment III except that the photographs showed only the subject's behavior.

Subjects. A new group of thirty-five college freshmen served as judges.

Results. Table 2 shows that the accuracy for the total trials and for A photographs was significant, while the accuracy for B photographs was not significant. Comparison of the results of Experiments III and IV shows that there was no difference for A, while for B there was a significant difference ($p<.01$).

DISCUSSION OF EXPERIMENTS I-IV

When accuracy across all the photographs from interviews A and B is considered, three of the four experiments provide evidence of significant accuracy in identifying the stress and catharsis phases of the interviews. The results are not nearly so uniform, however, when we consider the accuracy achieved for each of the two interviews. It is clear that these four experiments alone raise more questions than they conclusively answer, demanding additional experiments in which judges respond to

photographs from new standard interviews. The emphasis of the discussion of this first set of experiments will be on these questions and the decisions they prompted for the design of additional experiments.

The accuracy level achieved seemed to have been influenced by three variables: (1) the number of persons shown in the photographs, viz., either both interviewer and subject or only subject; (2) the particular stimulus persons shown, viz., those from A or from B; (3) the method of presenting the nonverbal behavior, viz., to elicit from each judge either a single response to a single photograph or a response to each of a number of photographs.

1. When only the subject was shown, significant accuracy was found for only one of the two subjects, and only when judges rated many pictures. In contrast, the greatest accuracy was achieved when both interviewer and subject were shown. It is difficult to specify the reasons for this, since the Interviewer-Subject condition may have contained at least two types of information not present in the Subject-alone condition, and the interviewer's role may have contaminated both of them. When a judge sees both interviewer and subject he may observe something about the interplay between two people not evident when he sees either interviewer or subject separately. In addition to such clues about the interaction, the interviewer displays his own set of nonverbal cues, and the judge may look at both interviewer and subject separately and summate in some fashion his impressions of each. Interviewer and subject differed, however, in their knowledge of the experiment, the degree of their involvement, and the nature of their role. The subject did not really know what the procedure was about; he did not know the focus of the research; and his behavior was reactive to that of the interviewer, and spontaneous in the sense that he did not follow a prearranged plan. The interviewer was playing a standardized role; he knew the focus of the research; he knew what he was trying to do to the subject and how he should proceed. In comparison to the subject's actions, the interviewer's may have been more stereotyped and artificial. Since judges did best when the interviewer was shown, we are confronted with the familiar problem of posed behavior—a problem these experiments were designed to avoid. It may be that accuracy was due more to the posed behavior of the interviewer than to the more spontaneous behavior of the subject. Since our primary interest is in the latter, the decision was made to show only the subject's behavior in further experiments.

2. When we consider the results for the Subject-alone condition, a trend is evident toward better results on A than on B in Experiment II, with the difference significant in Experiment IV. It might be that the two subjects differed in their nonverbal expressiveness, that one sent more information than the other through the nonverbal channel. Intriguing as it might be to raise questions about individual differences in nonverbal communicativeness, a much simpler explanation may account for the differences in the results. Perhaps persons A and B are equally expressive in their nonverbal behavior, but their experience in the interview differed.

For example, if person A fully experienced both stress and catharsis, but person B though stressed did not experience catharsis but found the whole interview unpleasant, then judges would regard person B's "catharsis" photographs as stressful; and we, the experimenters, would call the judges wrong.

Further exploration of individual differences in nonverbal communicativeness must therefore better insure that the participants have a comparable experience. One step in this direction is to utilize the same interviewer, so that even if the subjects differ in their mode of handling the experience at least they encounter the same person.

The decision was made, therefore, to employ only one interviewer in the next series of experiments. (The consideration of whether differences in subject experience confuse a measure of nonverbal communicativeness will be presented in a later section, and a different method of measuring nonverbal sending will be offered as another solution.)

Furthermore, the new experiments were planned to permit a larger sampling of photographs than was possible with A, and a comparable number from each interview. In order to permit focus on individual differences between senders, accuracy will be reported only for judgments of each stimulus person.

3. The final decision made was to utilize the multiple (rather than single) judgment procedure in order to make more economical use of a judge population, and to build a pool of judge reactions to as many of the specific photographs from each interview as possible. Plans were made to check for benefits due to practice; data on practice effects in these experiments will be reported later, with analysis of practice effects on the subsequent experiments.

EXPERIMENTS V-IX

Before proceeding to the experiments with Interviews C, D, and E, one more experiment relevant to A and B was needed to explore and, if possible, refute the notion that the results were derived from a factor extraneous to our fundamental hypothesis. The task for the judges in the first four experiments was specified as identifying the stress and catharsis phases of an interview from nonverbal cues. Since stress had always preceded catharsis, it might be argued that judge accuracy was not based on the linking of nonverbal cues to the stress and catharsis phases of the interview as such, but more simply on recognition of nonverbal behavior typical of the start or end of any interview. Accordingly, Experiment V was conducted to determine whether judges could accurately identify photographs taken at the beginning and end of an interview if they were told nothing about the nature of the interaction. The procedure was identical to that of Experiment IV except that the instructions omitted any mention of the stress and catharsis phases, and instead stated that in a standard interview the interviewer had acted one way at the start and a different way at the end. Judges were asked to decide

whether photographs were taken in the first or second part of the interview. The results, shown in Table 2, indicate that accuracy was not significant, and demonstrate that the accuracy in Experiment IV was dependent at least in part upon the judges' cognizance of the stress-catharsis interview relationship.

Experiment VI

Problem. The purpose of this experiment was to determine whether, with a larger sampling of nonverbal behavior from two new interviews, results from the earlier experiments would be replicated to provide further evidence that nonverbal behavior communicates accurate information about two interview conditions. Hypothesis: Untrained judges can accurately determine whether randomly selected and ordered photographs were taken during the stress or catharsis phase of standard interviews.

Method. Twenty photographs were randomly selected from the stress phase and 20 from the catharsis phase of Interview C; the same was done for Interview D. The photographs from each interview were arranged in a random sequence. Judges were shown all 40 photographs of one subject, and then the 40 photographs of the other subject. Judges were randomly assigned to two groups; group one responded first to person C and then to person D, while group two had the reverse order. In all other respects the procedure was identical to that of Experiment IV.

Subjects. A new group of 49 college freshmen served as judges.

Results. Since there were no differences in the accuracy scores achieved by judges who had seen person C first and those who had seen person D first, the scores for the two groups were combined. The obtained accuracy scores reached significance for both C photographs and D photographs. Table 3 shows, however, that the median number of pictures of person C accurately identified was only slightly above .50 correct. The photographs of person D were identified with significantly greater accuracy than those of person C ($x^2=16.41$, $p<.001$). The distribution of scores was skewed towards accuracy for person D.

Discussion. These results closely replicate the findings reported for Experiment IV, despite the difference in persons shown in the photographs and the use of new groups of judges. It might be noted, further, that a year elapsed between the first four experiments and Experiment VI; thus, the judge population and other time-related factors could vary.

Once again, as in Experiment IV, there is a difference in the level of judge accuracy on the two persons. The accuracy level for C is almost the same as for B, while A and D seem to elicit a higher level of accuracy. Examination of the interview typescript and inspection of the photographs suggested that the subject in C experienced less catharsis and was more uncomfortable throughout the interview than the subject in D. Thus, again, the question of individual differences arose. In the discussion of the results for all the experiments, this question will be explored.

Table 3 Results of Experiments VI, VII, VIII, and IX
Multiple judgment procedure

Pictures correctly identified	Experiment VI (49 Judges) 40 Photos of C 40 Photos of D		Experiment VII (70 Judges) 36 Photos of D 36 Photos of E		Experiment VIII (40 Judges) 60 Photos of D 60 Photos of E		Experiment IX (74 Judges) 100 Photos of E	
	Score	Proportion	Score	Proportion	Score	Proportion	Score	Proportion
Interview C								
Median	21.0*	.52	- - -	- - -	- - -	- - -	- - -	- - -
1st quartile	18.8	.47	- - -	- - -	- - -	- - -	- - -	- - -
3rd quartile	22.8	.57	- - -	- - -	- - -	- - -	- - -	- - -
Interview D								
Median	25.7***	.64	21.9***	.61	39.2***	.65	- - -	- - -
1st quartile	23.0	.58	19.6	.54	35.8	.60	- - -	- - -
3rd quartile	27.2	.68	23.8	.66	43.0	.72	- - -	- - -
Interview E								
Median	- - -	- - -	21.8***	.61	38.8***	.65	68.0***	.68
1st quartile	- - -	- - -	20.9	.58	36.5	.61	63.4	.63
3rd quartile	- - -	- - -	23.6	.66	41.5	.69	72.1	.72

*$p < .05$
***$p < .001$

Experiment VII

Problem. The purpose of this experiment was to replicate the findings on D and extend the findings to another subject, the subject in E. Hypothesis: Untrained judges can accurately determine whether a randomly selected and ordered series of photographs were taken during the stress or catharsis phase of an interview.

Method. Thirty-six photographs were randomly selected from D, 36 from E. The procedure was identical to that of Experiment VI.

Subjects. A new group of 70 college freshmen served as judges; half viewed E first, then D; half followed the reverse order.

Results. Again there was no significant difference between those judges who had seen E first and those who had seen D first, and the scores for the two groups of judges were combined. The obtained accuracy scores were significant for person D and for person E photographs, and there was no difference in accuracy on the two subjects. Table 3 presents these results.

Discussion. The results on person D replicate the earlier results on this person; the results on person E are quite similar to those reported for persons D and A.

Experiment VIII

Problem. This experiment was designed to explore judge ability to interpret the behavior of different persons, a problem to be discussed in a later section. It is reported here only as a replication of Experiment VII. Hypothesis: Untrained judges can accurately determine whether randomly selected and ordered photographs were taken from the stress or the catharsis phase of an interview.

Method. The procedure was identical to that of Experiment VII except that a new group of photographs was randomly selected from D and E, and 60 pictures, rather than 36, were chosen from each interview.

Subjects. A new group of 40 night-school students served as judges.

Results. Again there was no difference in the accuracy scores of those who saw E first and those who saw D first, and the results were combined. The obtained accuracy was significant for photographs of D and E. Table 3 shows these results.

Discussion. These results closely replicate the earlier findings. The findings on D have been quite consistent across the different independent groups of judges, and for different randomly selected samples of photographs in Experiments VI, VII, and VIII. The same is true for E in Experiments VII and VIII.

Experiment IX

Problem. This experiment addressed the problem of stability of judgment of nonverbal behavior over time. The problem will be discussed later, but part of the results are reported here as a final replication of this series of experiments. Hypothesis: Untrained judges can accurately determine whether randomly selected and ordered photographs were taken during the stress or the catharsis phase of an interview.

Method. The procedure was identical to that of Experiment VIII except that only photographs of E were shown. A new sample of 100 photographs was randomly selected, and judgments were made twice by each judge, with a four-day interval between trials. Two different randomly determined orders of presentation were employed.

Subjects. A new group of 74 college freshmen served as judges.

Results. The results for the first testing are shown in Table 3. Again, E was accurately judged, replicating the results of Experiments VII and VIII.

Discussion. This experiment replicated the accuracy results on E. That is, the same general level of accuracy was achieved on E in Experiments VII, VIII, and IX. It should be noted that for each experiment a different sample of photographs was randomly selected from the interviews, and a different group of students served as judges.

These experiments did vary in the absolute number of photographs shown to judges, and it appears that the median number of photographs correctly identified slightly increased as the size of the photograph sample increased. For person E the increase was as follows:

in Experiment VII the 36 photographs gave a median accuracy of .61;
in Experiment VIII the 60 photographs gave a median accuracy of .65;
in Experiment IX the 100 photographs gave a median accuracy of .68.
The increase was as follows for person D:
in Experiment VII the 36 photographs gave a median accuracy of .61;
in Experiment VI the 40 photographs gave a median accuracy of .64;
in Experiment VIII the 60 photographs gave a median accuracy of .65.
While the increases in accuracy are small, they show at least that any loss which might be due to fatigue associated with making an increased number of judgments is more than counteracted by the benefits associated with seeing a larger sample of interview behavior. The benefits, practice effects and the opportunity to acquire ideas about the subjects' range of behavior, made possible by the use of many photographs, will be discussed next.

FURTHER ANALYSIS OF EXPERIMENTS III-IX

The data from Experiments III-IX were further analyzed to examine the influence of practice on accuracy, and the ability of a judge to achieve significant accuracy for more than one stimulus person (interview subject).

Practice Effects

With the use of the procedure in which judges gave separate responses to many photographs, in Experiments III-IX, it seemed reasonable to expect that there might be some increments in accuracy over trials as a result of practice in performing the judgment task. Such a possibility was explored by comparing a judge's accuracy on the first third of the trials with his accuracy on the next third of the trials. The judgments of the last third were not included, since any benefits due to practice might possibly have been counteracted by fatigue or boredom toward the end of the procedure. A Wilcoxon matched-pairs signed-ranks test was performed on the difference for each judge between his first-third and second-third accuracy scores, employing a two-tailed significance test. Table 4 shows these data; results are not given for person A, because fourteen trials did not provide enough judgments for measuring practice effects, nor for Experiment V, because in that experiment the judgment task concerned interview chronology rather than the stress and catharsis phases.

The results were inconsistent. In some experiments the increase in accuracy scores was not only statistically significant but substantial in size; for example, in Experiment III the increase in accuracy on B was 20 proportion points, and in Experiment VII the increase in accuracy on E was ten proportion points. On the other hand, significant decreases in accuracy also occurred, even with the same stimulus person in different experiments; for both persons D and E, accuracy significantly increased in one experiment, significantly decreased in another experiment; and there was the further inconsistency that in the same experiment in which

Table 4 Practice Effects

Experiment	Stimulus persons	Number of trials in each third	Median first third trials	Median second third trials	p
III	B	11	4.85	7.05	< .001
IV	B	11	4.81	5.85	< .05
VI	C	13	6.45	5.55	> .10
	D	13	8.88	7.88	< .05
VII	D	12	7.00	6.85	> .10
	E	12	6.00	7.14	< .001
VIII	D	20	11.80	14.50	< .001
	E	20	12.14	11.10	< .05
IX	E	30	19.00	19.50	< .05

judgments of D increased, judgments of E decreased. The inconsistency of these results on possible practice effects is in marked contrast to the consistency of the accuracy scores across all trials noted for the different stimulus persons across the various experiments.

Our interpretation of this confusing state of affairs is that, while practice may have an effect, the effect is probably contaminated by differences in the difficulty of judging particular photographs. It should be remembered that a different random selection of photographs of each stimulus person was used in each of Experiments VI-IX. Thus, the inconsistent results for a given stimulus person may be due to differences in the degree of difficulty represented by the particular selection and ordering of two sets of photographs; in one experiment the first third might have been more difficult to judge, while in another experiment the second third might have been more difficult. The results on person B were consistent in Experiments III and IV, in which the same photographs in the same order were shown.

A check was made to determine whether an equal number of stress and catharsis photographs had been presented in the first-third and second-third of the trials in the various experiments, since a differential rate of guessing either response might have led to greater or lesser accuracy if the two types of photographs were unequally distributed. But they were not. Further, practice effects were again measured for each experiment to compare the first ten trials with the last ten trials, and almost identical results to those reported in Table 4 were obtained. Thus, we conclude that it was not possible to determine whether an increase in accuracy results from practice over trials.

Judge Ability Across Stimulus Persons

These experiments were not designed for studying the characteristics of the good and bad judge of nonverbal behavior, nor were they chiefly intended to demonstrate that judging ability is consistent across different stimulus persons. This latter point, however, does raise some relevant questions about the nature of nonverbal communication. How much overlap is there in the repertoire of nonverbal cues—specific facial expressions or body positions—of two different persons? Does the same nonverbal cue have the same meaning in the repertoire of two different persons?

A procedure for establishing similarities in accuracy for a given judge across different stimulus persons (interview subjects) would be to correlate his accuracy scores on two stimulus persons. There are three possible outcomes, supporting one or more interpretations. If similar nonverbal cues are displayed by different stimulus persons and have the same meaning for both persons, then a judge who can decipher one person's nonverbal behavior could do so with another's, and the correlation coefficient for judge accuracy on any two people should be positive. If the nonverbal cues shown by two stimulus persons are dissimilar, but a judge is familiar with the meanings of the two distinct sets of cues, then again he would score similarly on both persons, and the correlation coefficient would again be positive. If two sets of nonverbal cues are dissimilar and the judge is familiar with only one of the types, then positive or negative results on one stimulus person would not predict a judge's ability to interpret the behavior of another, and the correlation would not be significant. If different stimulus persons display some of the same nonverbal cues, but these cues have different meanings for the different stimulus persons, then a judge who applies the same standards of interpretation to a given cue whenever it appears will be correct for one person and incorrect for another person, and the correlation will be negative.

This is not meant to be an exhaustive list, but rather an indication of some of the questions which could be raised by a study of judge ability across different stimulus persons. Unfortunately, as we shall soon see, the data from the present experiments do not very well lend themselves to this type of analysis.

The correlations of judges' accuracy on different stimulus persons is reported in Table 5 for Experiments III, IV, VI, VII, and VIII. Before Experiment VIII was designed, the distribution of accuracy scores in the other experiments was examined to explore the reasons for the low correlations. These distributions of accuracy scores are quite skewed and limited in range; almost all of the judges scored correctly on between .55 and .75 of the photographs. Such a limit on the distribution of scores would of course greatly restrict the size of any correlation coefficient between two sets of accuracy scores. Experiment VIII was designed to overcome this limitation. A larger sample of photographs of each interview subject was shown, and the judge population, night-school students, was selected

because of their greater heterogeneity as compared to daytime students. The median age for the judges in Experiments I-VII was 21.5, with an inter-quartile range of less than two years; the median age for the judges in Experiment VIII was 31.3, with an inter-quartile range of about seven years. These efforts to enhance the correlation coefficient seem to have been modestly successful; the largest correlation coefficient achieved was in this experiment.

Table 5 Correlations between accuracy scores on different Ss

Experiment	Number of judges	Stimulus persons shown	Rank order correlation
III	33	A & B	.319*
IV	35	A & B	.233
VI	49	C & D	.269
VII	70	D & E	.217
VIII	40	D & E	.44**

*p < .05
**p < .01

The data from these nine experiments suggest that there is a positive correlation in judging ability across stimulus persons. The questions outlined earlier would be more meaningfully studied with a test which not only yields a greater range of accuracy scores, but, more importantly, which presents the nonverbal behavior of stimulus persons systematically selected for their differences in personal characteristics. Work presently in progress is exploring judges' reactions to photographs of different psychiatric groups.

SUMMARY OF EXPERIMENTS IV-IX

Experiments IV-IX differed from the earlier experiments in some features of their design. Experiments I and III showed both interviewer and subject; Experiments I and II required only one judgment from a judge for one photograph. Experiments IV-IX all showed only the subject, and all required a judgment for each of many photographs from each judge.

Summary of method. In Experiments IV-IX judges in every case saw a number of randomly selected and ordered photographs of each stimulus person they judged, and in every case saw only the subject without the interviewer. Viewing time for each photograph was twenty seconds, during which time judges wrote their decision as to whether the picture had been taken during the stress or catharsis phase of the interview which had been described generally in the instructions. The actual number of photographs shown varied for the five interviews from 14 to

100. Across all of the experiments judgments were obtained on a total of 14 photographs of A, 34 photographs of B, 40 of C, 100 of D, and 120 of E. Photographs were selected in every case in a random fashion without prior inspection of the pictures; in most of the experiments a new random sample was selected each time a particular interview was to be shown.

Summary of judge population. Independent groups of judges viewed the photographs. In every experiment except one these were freshmen psychology students; in Experiment VIII a more heterogeneous group of night-school psychology students served as judges. In all, 35 judges responded to photographs of A, 35 judged B, 49 judged C, 159 judged D, and 184 judged E.

Summary of results. Significant accuracy was achieved for at least one of the stimulus persons judged in all of the experiments except Experiment V, which was designed as a control experiment and was not expected to produce accuracy. Significant accuracy was achieved on four of the five stimulus persons judged, although for one of the four, person C, the level of accuracy was just above the chance level. The levels of accuracy were quite similar on persons A, D, and E; the median proportion of photographs accurately judged ranged from .69 on A, to .61, .64, and .65 on D, and .61, .65, and .68 on E. The results on persons D and E were replicated across three independent groups of judges, who in each case responded to a different random sample of photographs.

The accuracy results were similar for B and C, and differed from the three interviews just described in that either slight but significant or nonsignificant accuracy levels were obtained. The median proportion of photographs correctly identified was .51 for B and .52 for C.

Practice effects were evaluated; results were inconclusive.

In two of the nine experiments a significant low-to-moderate correlation in judge accuracy among the different stimulus persons was obtained. In the other experiments the correlations were also positive, but not significant.

GENERAL DISCUSSION

With the establishment of the proposition that nonverbal behavior as shown in still photographs provides accurate information to observers about the two phases of a standard interview, many questions are raised: (A) How much of the nonverbal behavior shown during an interview actually communicates information? (B) What kind of information does nonverbal behavior provide? (C) What might account for the differences in judge accuracy between stimulus persons; or, do differences in judge accuracy reflect differences in the nonverbal communicativeness of individual stimulus persons, and if so, what do such differences mean? (In discussing these questions data from two new series of experiments will be introduced.)

A. THE AMOUNT OF NONVERBAL BEHAVIOR COMMUNICATING INFORMATION

Most generally phrased, our concern is with how rich nonverbal behavior is as a source of information. But, there are two aspects to this question: *what different kinds* of information may be learned from a given sample of nonverbal behavior, and *how much* of the behavior sampled actually communicates systematic information of any kind. This latter question will be discussed first, and can be stated operationally as *how many* of the photographs provided information relevant to the judgment task. In answering this question we will draw a distinction between photographs which elicited from the judges substantial agreement, regardless of accuracy, and photographs about which judges usually disagreed. The former will be defined as pictures which provide *consistent* information, and the latter will be considered as *ambiguous* photographs.

In deciding how much of the nonverbal behavior communicated information we will consider the following: (1) the ratio of consistent to ambiguous photographs; (2) the meaning of inaccuracy; (3) a new experiment which studied a further question to be raised in the second section.

1. Do most of the photographs convey consistent or ambiguous information?

The levels of accuracy in the first nine experiments can be cited in support of either of two contradictory hypotheses: that most of the photographs were ambiguous, and that most of the photographs communicated consistent information.

On the hypothesis that most of the pictures were ambiguous, it might be argued that, since the median proportion of correct indentification never reached .75, judges were responding in a chance fashion to *most* of the pictures and correctly identifying only those few which do carry clear information, thus raising the median accuracy level from .50 to .60-.70. On the hypothesis that most of the pictures carried consistent information, the moderate level of the median proportion correct can be attributed to most of the photographs carrying either *consistently inaccurate* or *consistently accurate* information, thus canceling each other out, in part.

In the discussion to follow, a *consistent photograph* is one to which the majority of judges gave the *same* designation, whether "stress" or "final," whether accurately or inaccurately in terms of the task. An *ambiguous photograph* is one about which there was considerable disagreement. Boundaries were arbitrarily set, as follows: consistent photographs were those which were either correctly identified or incorrectly identified by more than 60% of the judges. Consistently accurate photographs were those which more than 60% of the judges correctly identified; consistently inaccurate photographs were those which less than 40% of the judges correctly identified. The boundaries of the criterion for ambiguity were set at ten percentage points on either side of .50, around the point at which half the responses to a given picture were "stress," and

half were "final." Figure 2 shows examples of consistent and ambiguous photographs.

The data from Experiments IV-IX were further analyzed in order to provide a basis for choosing between the two hypotheses as to whether

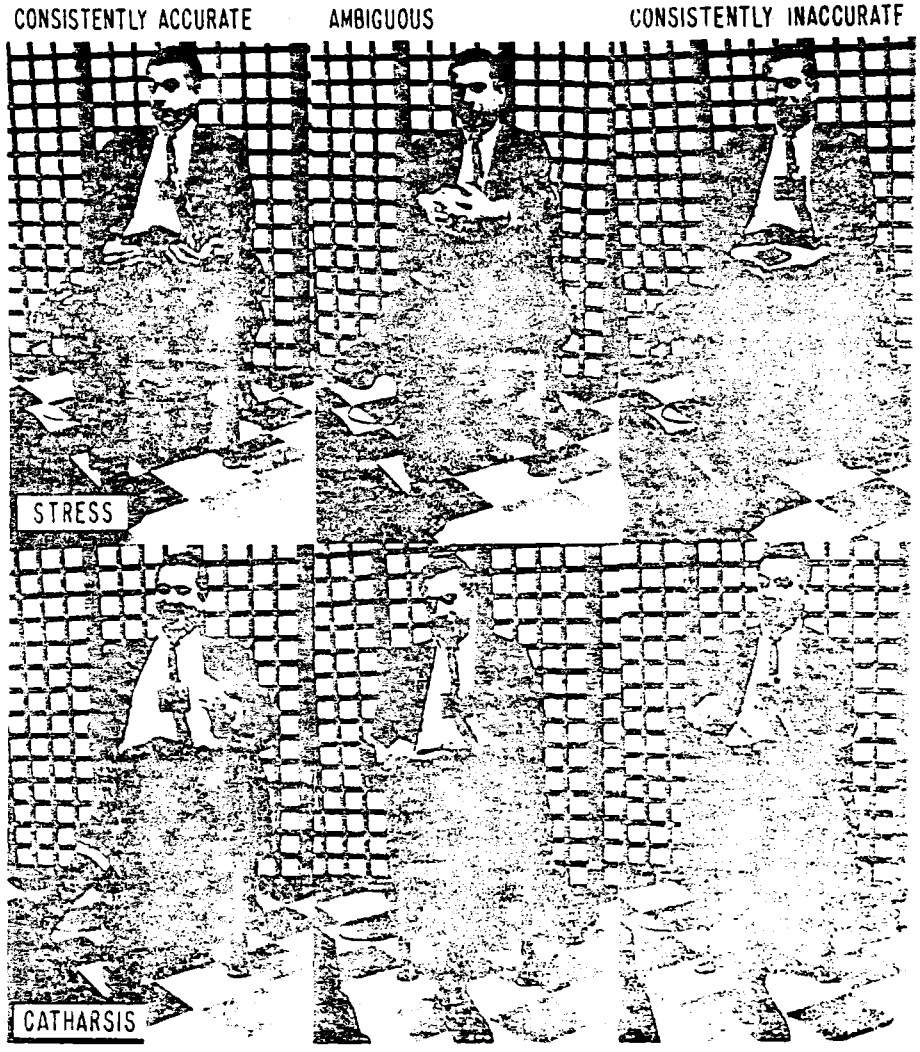


Figure 2. Examples of consistent and ambiguous photographs.

most of the photographs communicated consistent or ambiguous information. The first analysis compared the responses of different groups of judges to the same sets of photographs to determine through a rank

order correlation whether the photographs maintained their same relative position in terms of the proportion of judges accurate on each picture. The judges who had viewed the photographs in Experiments IV, VI, VII and VIII were randomly assigned to two subgroups; the proportion of correct responses to each photograph was calculated for each subgroup and a Spearman rank-order correlation (Siegel, 1956, p. 202) was performed. The correlations, shown in Table 6, are generally high, indicating that the photographs tend to remain in the same relative position in terms of accuracy of identification when viewed by different groups of judges. These results are interpreted as providing some support for the hypothesis that most of the photographs carry consistent information. If most of the photographs were in fact ambiguous, certainly it would be unlikely that they would maintain the same relative standing when judged by different groups.

Table 6 Rank order correlation between proportion of correct responses to each photo

Experiment	Interview	Number of judges within each group	Rank order correlation by photo
IV	A	17 and 18	.81
	B	17 and 18	.87
VI	C	25 and 24	.73
	D	25 and 24	.84
VII	D	35 and 35	.86
	E	35 and 35	.88
VIII	D	20 and 20	.63
	E	20 and 20	.85
IX	E	100 test-retest	.92

All of the correlations are significant beyond .001 level of confidence.

The data from Experiment IX allowed examination of another type of consistency: consistency over time for the same set of judges. In Experiment IX the same group of judges viewed the same pictures on two occasions separated by four days. A Spearman rank-order correlation was performed between the numbers of correct responses to each photograph on the two time occasions. This correlation is higher than the others reported in Table 6, probably because this was the only instance in which the same judges viewed the same pictures.[3] This correlation was interpreted as also supporting the hypothesis that most of the pictures provide consistent information. The reasoning is that, if most of the pictures were ambiguous, they could not be expected to maintain so well their relative

[3] Although less relevant to the problem under discussion, it might be mentioned that Experiment IX also offered the opportunity to establish whether judges maintained the same relative level of accuracy over time. Thus, a rank order correlation was also computed on the basis of the number of photographs accurately identified by each judge on the two time occasions. This Rho was .60, $p<.001$, showing a moderate stability on judge performance.

position when judged on two different occasions by the same judges.

Another approach to the problem was to examine the distribution of scores for each interview to find the numbers of photographs judged consistently accurately, consistently inaccurately, and ambiguously, according to the criterion previously described. The number of judges correct on each photograph was tabulated. The proportions of correct identifications were calculated by totaling the correct judgments for each photograph across all experiments in which it had been viewed and dividing the sum by the total number of judges who had seen the picture. A frequency distribution was drawn, grouping these proportions of correct identifications into ten-point intervals. The frequency of photographs within each interval was then converted into a proportion of the total number of photographs for which judgments had been obtained; this step was necessary to adjust for the considerable divergence in the

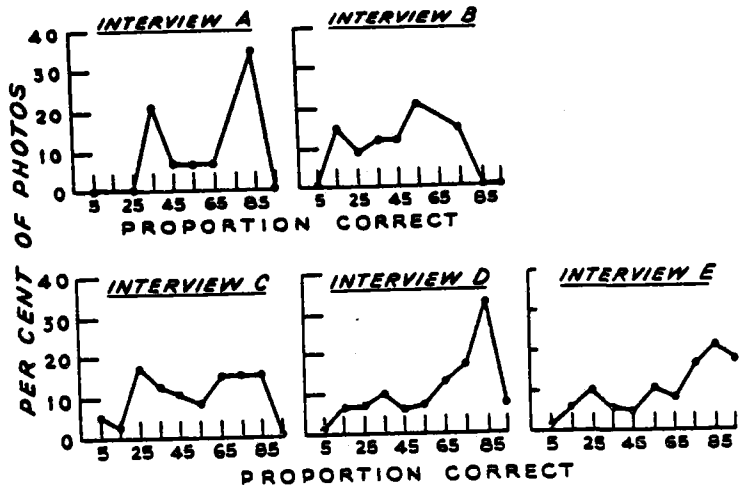


Figure 3. Distribution of photographs within each interview in terms of the proportion of judges who correctly identified each of the pictures.

numbers of photographs judged from each interview. Figure 3 shows these data, which bear directly on the choice between the two hypotheses under discussion. If most of the photographs were ambiguous, we should expect distributions symmetrically centered around the .50 level of correct identification, with the greatest concentration near the .50 level. Clearly, this is not the case. The figure shows that some photographs were inaccurately judged (below .40), some were in the ambiguous range, and many were accurately judged (above .60), thus supporting the hypothesis that most of the pictures provided consistent information. These data have been summarized in Table 7. (From both Table 7 and Figure 3

Table 7 Ambiguity - Consistency

Interview	Number of photos for each interview	Ambiguous photos (41 to 60% of Js correct)	Consistent Photos		
			Accurate (61 to 100% of Js correct)	Inaccurate (0 to 40% of Js correct)	Accurate and inaccurate
		%	%	%	%
A	14	14	65	21	86
B	34	32	32	36	68
C	40	17	45	38	83
D	100	11	67	22	89
E	120	15	63	22	85

it is evident that there were differences among the stimulus persons in the five interviews; these data will be further examined later in the discussion.)

2. What does inaccuracy mean?

In the course of establishing that the majority of the photographs communicated consistent information, we have discovered that some of the pictures in each interview were consistently inaccurately interpreted. Now, the question is, how do we interpret this phenomenon?

The accuracy criterion was simply that judges correctly identify the phase when a picture was taken. Accuracy, therefore, depended in some part on the success of the standard interview procedure, on the interviewer's eliciting two quite distinct reactions from the subject during the two phases.

Even if we assume such success, inaccuracy could still occur in two ways. Particular expressions reflective of a phase for a particular stimulus person might be peculiar to him, and in fact might connote just the opposite experience to most observers. For example, a slouching posture shown during stress might be a habitual uncomfortable response for a given stimulus person, and yet be usually, but inaccurately by our criterion, judged as a cathartic response. Another possibility is that at certain moments a subject may have tried to disguise his feelings or conceal them from the interviewer, and his masking cues might have misled the judges. For example, a subject's hands carefully folded in his lap during stress might have been read by judges as a poised or relaxed nonverbal cue, and incorrectly called catharsis. Thus, both possible idiosyncracies in the subject and possible attempts to mask his actual experience could account for inaccuracy.

If, to these sources of possible confusion arising from the subject, we add the possibility that the interview procedure was not successful in eliciting two quite distinct reactions in the subjects, or in all subjects, then we must consider a different kind of inaccuracy—one which is at-

tributable to a limitation in the experimental control of the interview. Discussions with the participants after the interview, their responses to their own photographs, and our own reading of the typescripts, all suggest that the standard interview format was not completely successful. It appears that the subject's experience was not unidimensional within each phase. For example, it seems likely that there were moments during the stress phase when the subject felt he had temporarily avoided attack, or obtained the upper hand, and that these experiences as captured by the photographs appeared to judges to be part of the catharsis phase. Similarly, moments during the catharsis phase when the subject introspected about his behavior during stress, or wondered why the interview was still continuing, might have been stressful for him, and been regarded by judges as part of the stress phase.

It is doubtful that experimental manipulations of a standard stress interview can completely control for a subject's repertoire of defensive tactics to meet the interviewer's assault, or for the interviewer's failure in a particular line of inquiry to evoke a particular feeling in a subject. We may conclude that the inaccuracy results reflect, at least in part, this lack of complete experimental control over the subject's experience. Therefore, we must not interpret inaccuracy as evidence that the nonverbal behavior necessarily provides wrong or misleading information.

3. *Does accuracy increase if judgments are based on more nonverbal cues?*

It is possible that accuracy might have been substantially improved if short bursts of motion picture film had been shown instead of still photographs, so that judges could observe movement in addition to positions, and stimuli more closely resembling behavior ordinarily seen. In a recently completed doctoral dissertation on the training of judges to interpret nonverbal behavior, Hoffman (1964) used as his stimuli five-second bursts of motion picture film. He employed a stress interview procedure patterned after ours, with four different subjects and the same judgment task of guessing whether each stimulus unit was from the stress or catharsis phase of the interview. Thus, our experiments are comparable except that Hoffman's showed 24 frames per second for five seconds and ours showed one frame. The accuracy obtained by Hoffman's judges was remarkably similar to that reported here; prior to training, the average accuracy across all judges across all four stimulus persons was .56 of the photographs correctly identified, with a range of average accuracy of .48 to .64. Clearly, for this judgment task short motion picture film bursts yield no greater accuracy than the still photograph.

Another feature of both designs, however, might account for our obtaining only moderate accuracy. Judges were forced to respond to behavior presented completely out of context. Possibly, inaccurate judgments could be avoided with an experimental design which shows to judges a larger parcel of consecutive behavior, thus permitting them to disregard a set of cues which are inconsistent with the general tenor of the experience. Certainly, in usual interactions many more nonverbal cues are

available to observers than are shown in either a split-second photograph *or* five seconds of motion picture film shown out of a larger temporal context. More typically, a given nonverbal cue is seen in the context of behavior which precedes and follows it, and observers can evaluate the behavior within that framework. An experiment was therefore designed which differed from the previous nine in that it provided judges with a substantially increased and sequential sample of nonverbal behavior.

Experiment X

Problem. The purpose of this experiment was to determine whether the addition of sequential context cues would increase accurate judgment. Hypothesis: Providing judges with all of the photographs from a single interview phase will lead to higher accuracy levels than have previously been found.

Method. The instructions and procedure were the same as those of the earlier experiments with the following exceptions. Judges were shown all of the photographs from a single phase of the interview in succession, with each picture projected for five seconds. After seeing all of the photographs from a phase, judges recorded their decision—viz., whether the pictures were of the stress or catharsis phase. When making their judgment they did not know that they would next see all of the photographs from the other phase. Before seeing the other set of photographs they were told that, when they had seen the second set, they would again judge the *first* set of pictures. They were specifically instructed to write down a second decision about the *first* series of photographs, and not to change their first answer; moreover, they were told that there was no need to be consistent, that their second answer could be the same as or different from their first answer.

Photographs from D and E were shown because the earlier studies produced good results with these interviews, and because fairly extensive photographic records of each were available: 80 photographs were shown of person D during stress, 80 during catharsis; 74 photographs were shown of person E during stress, 74 during catharsis. There were four experimental groups; half saw D, half saw E; half saw the stress phase first, half the catharsis phase.

Judges. Undergraduates from upper division psychology courses served as judges. Since this experiment was conducted with summer school students, the range of age was greater than in any of the judge samples reported earlier; in order to obtain a more comparable age group, persons over 30 years of age were discarded from the data analysis. Forty-four judges remained of D, 32 of E.

Results. Table 8 shows the results for the first and second decisions. The hypothesis pertained only to the first decision, for which the judges had seen all of the photographs from an entire interview phase, but only from one phase. The prediction was that the accuracy level would be

Table 8 Results of Experiment X

Interview	Number of Js	Number (and proproportion) of Js accurate first decision	second decision
D	44	28 (.636)	38 (.864)
E	32	19 (.594)	23 (.719)

greater than was found in the previous experiments where no sequential cues were available. There had been considerable variation in the design of the first nine experiments, however, and none of them allows an exact comparison in which all variables could be held constant except for that of sequential cues. Table 9 compares the experiments on six aspects of their design.[4]

Experiment II is comparable with Experiment X in that both showed only the subject, required only one judgment from each judge, and

[4] Although not directly relevant to the problem under discussion, the data from Experiment X can be analyzed to yield some ideas about comparative accuracy on the two phases of the interview. The reader may have wondered why there has been no such presentation of results for the stress and catharsis photographs separately in the earlier experiments. The reason is that such an analysis would be quite unsound, since the judges' rate of guessing either stress or catharsis would confuse the interpretation of such results. If, for example, a judge guessed stress most of the time, he would obtain higher accuracy on stress than on catharsis photographs, although, it should be noted, that such a guess rate would not contaminate the accuracy measure across all photographs. Inspection of the data from the first nine experiments did reveal that most judges guessed stress more frequently than catharsis; whether this is because the pictures are more stressful than cathartic, or because of response set is unknown. Experiment X is not entirely free of this problem, but at least judges have responded to only one set of photographs, either stress or catharsis. These data have been relegated to a footnote, however, since the judges' stress guess rate could still cause a difficulty in interpretation, and if our interest had been to study the differences between the two phases a more elaborate design would have been necessary in which the judges' guess rate was first assessed, and the subjects matched on this basis.

The differences in the accuracy on the two phases in Experiment X, in any case are not particularly impressive. For Interview D, .70 of the judges were accurate on stress, .58 on catharsis; for Interview E, .59 were accurate on stress, .60 on cartharsis. These results were found on the first decision made in Experiment X, and it appears that there is a trend toward more accurate judgment of stress for Interview D, but not for Interview E. The results found for the second decision are .95 of the judges correct on stress, .79 correct for cartharsis on Interview D; .76 correct on stress, .57 correct for catharsis on Interview E. While stress appears to be more correctly identified by more judges on both D and E, interpretation is further confounded by the fact that when they made the second decision judges had seen both phases of the interview, and differential accuracy may be due to order effects, willingness to change one's mind, etc. These data have been presented more to satisfy any curiosity about why stress and catharsis have not been separately analyzed than because of the substantive value of these results.

Table 9 Features of the designs of the experiments

Experiment	Stimulus persons shown in photos	Interviews judged	Number of photos shown of each stimulus person to each judge	Judgments Pictures	Each judge sees photos from one or both interview phases	Absolute number of judgments by each judge
I	Interviewer & S	A & B	1	1/1	One phase	1
II	S only	A & B	1	1/1	One phase	1
III	Interviewer & S	A & B	14 A 34 B	1/1	Both phases	48
IV	S only	A & B	14 A 34 B	1/1	Both phases	48
VI	S only	C & D	40 C 40 D	1/1	Both phases	80
VII	S only	D & E	36 D 36 E	1/1	Both phases	72
VIII	S only	D & E	60 D 60 E	1/1	Both phases	120
IX	S only	E	100 E	1/1	Both phases	100
X (First decision)	S only	D & E	80 D 74 E	1/80 1/74	One phase	1

showed each judge only one phase of the interview; the two differed in that Experiment X provided not only sequential cues but also a larger number of pictures and a different interview sample. The results for Experiment II show that .58 of the judges of A and .47 of the judges of B correctly identified the interview phase, as compared to Experiment X, in which .64 of the judges of D and .59 of the judges of E correctly identified the interview phase. This difference in accuracy is misleading, however. It reflects the generally low accuracy of the judgments of person B. Since Experiment X had excluded persons who had been poorly judged (viz., person C), a more appropriate comparison would also exclude the results on person B. If only the results on person A from Experiment II are compared with the results from Experiment X, the accuracy difference is no longer very great.

Experiment VI-IX are comparable with Experiment X in that the same stimulus persons were shown, and many photographs were shown of each stimulus person; they differed in that Experiment X provided not only sequential cues but also showed only one phase to a judge and required only one judgment from each judge. The differences in accuracy are not striking. The median accuracy ranged from .61 to .65 of the photographs of person D correctly identified in Experiments VI, VII, and VIII, as compared to .64 in Experiment X. For person E, the median accuracy ranged from .61 to .68 in Experiments VII, VIII, and IX, as compared to .59 in Experiment X. Thus, the comparison of the results on the first decision in Experiment X with the results on the earlier experiments does *not* support the hypothesis that accuracy is greater when sequential cues are available to judges.

When the judges in Experiment X had seen the other phase of the interview, their second decision was better than their first. For D the accuracy level of the second decision, .86 judges correct, is considerably higher than that of the earlier experiments on D photographs.

Discussion. Common sense would suggest that adding more information, in particular providing clues about sequence of nonverbal behavior and its context, would increase accuracy; and yet this was not found to be so. The earlier discussion of the meaning of inaccuracy may help to explain this unexpected result. We interpreted the experiences of stress and catharsis as being not unidimensional, in that there was some overlap between the two phases which, when perceived by judges, led them to mis-identify certain pictures. This interpretation implies that nonverbal behavior may not only reflect an individual's gross affect changes, such as from stress to catharsis, but may also be fairly sensitive to and track the more moment-to-moment changes in an individual's handling of each experience. If the experience represented in the photographs is not unique or pure, then such momentary tracking may be faithfully reflecting the many features within the experience.

When a judge responds to many photographs singly he has no awareness of any of the contradictory elements within an interview phase; when he sees a smile during stress, for example, he does not know that it may have been preceded and/or followed by a terrible frown; he simply calls the photograph *cathartic* without any awareness of a problem—and his judgment is scored as inaccurate. When judges are shown an entire interview phase, they may benefit from context or sequence cues, but at the expense of having to contend with some of the seemingly contradictory messages sent within one interview phase. Admittedly, this is an ex post facto explanation, but it appears to be a reasonable one, and data from a new series of interviews lend some credence to it.

These new interviews were conducted with female patients hospitalized with a severe depression. Each patient was photographed at the time of admission to the hospital and again after some weeks of treatment shortly before discharge when she had made at least a partial recovery. Judges saw one entire interview and then were asked to decide whether the interview occurred when the patient was depressed or remitted. (The design was similar to that of Experiment X; but in these new experiments a judge saw only one interview and made one decision.) For this sample of depressed patients the mood experienced is quite uniform within an interview, the depression intense and durable, and so we could expect that judges would not be exposed to many contradictory or varied nonverbal messages within an interview. In the two experiments conducted to date, more than 90% of the judges correctly identified the interviews. The question of whether nonverbal behavior expresses momentary changes in an individual's mood will be discussed again in the following section about the kinds of information communicated by nonverbal behavior.

Experiment X does show, however, that when a judge sees the two contrasting sets of stimuli (both the stress and catharsis phases of the

interview) his second decision is considerably better than his first. Showing both phases of the interview, in addition to providing more comparative information, also informs judges that there was overlap between the two phases and thus may give a better basis for evaluating contradictions within each phase. (It can be noted that there was more improvement on D than on E; this difference is consistent with data from Experiment XI, to be presented later, which show that there was more similarity during the two phases in the behavior of person E than of person D.)

The reader may have some interest in the comparative value of a design like that of Experiment X, in which a judgment is based on seeing many photographs in sequence, and a design like that of Experiments III-IX, where judges respond to each of many single photographs shown out of order. While all of the experiments are artificial in that only nonverbal behavior is presented, judges are removed from the interaction, and still pictures rather than motion pictures are the basic stimuli, Experiment X is a little more akin to the more usual observation situation in that at least the behavior was shown in sequence, and the behavior sampled from an entire interview phase was presented. It is interesting to note that this advantage of being somewhat more true to life may be vitiated by our ignorance of the way in which people arrive at an interpretation of nonverbal behavior in ordinary interactions. That is, in an ordinary life situation, when persons observe a segment of nonverbal behavior we cannot know whether (1) the observer *utilizes all of the cues*, either by making definite judgments of each cue, summing them up and concluding "That man is stressed," or by making tentative partial interpretations, moment by moment, subject to modification by each succeeding cue; or (2) the observer *utilizes only a few cues*, either because others do not carry information or because the observer does not observe what information is there.

Showing a large amount of nonverbal behavior can be informative of how observers respond in more usual situations; some notions can be gained about the upper limits of accuracy, for example. But such a procedure does not help answer the question which has been discussed throughout this section of the paper. With a design like that of Experiment X it is impossible to resolve the question of how much of the nonverbal behavior carries consistent information; for if the observer responds to a mass of nonverbal behavior he could be picking up many separate cues from each photograph, all of which carry some information, or he could just as well be completely in the dark about most of the photographs and basing his judgment on a few consistently communicating pictures. Thus, while the procedure in Experiment X is more true to life, the procedure of obtaining judgments of more limited samples provides a better answer to the question of how much of the nonverbal behavior shown during an interview communicates consistent information. It should be noted that "the size" of the more limited sample, whether still photograph or a few seconds of motion picture film, is not the relevant issue. The critical point is that the design be one in which judges must

react to many separate slices of nonverbal behavior rather than to a conglomerate mass.

A second advantage of the design followed in the earlier experiments is that it allows study of the relationship between judgments and particular facial expressions and body positions shown. With the proposition now substantially established that judges respond systematically to nonverbal cues, it would be valuable to link judgments to particular nonverbal configurations, particularly if these apply to more than one stimulus person. Such analysis of these stress interviews and judgments is presently in progress.

B. THE KINDS OF INFORMATION COMMUNICATED BY NONVERBAL BEHAVIOR

In the preceding section we have seen that most of the nonverbal behavior shown during an interview communicates consistent information to observers. Our focus now shifts to an attempt to specify the types of information revealed by nonverbal behavior. Most generally, we are concerned with what an observer can learn from nonverbal behavior; specifically, our question is what were the kinds of information provided by the photographs which allowed judges to identify the stress and catharsis phases.

In most interpersonal settings, even those in which one person's role is defined as that of observer, nonverbal cues are not seen in isolation. The observer has verbal (content of a spoken message), vocal (voice quality), olfactory, and perhaps even tactile cues available concurrently with the nonverbal behavior (facial expressions, body movements and positions). The observer's knowledge of the situation, his expectations about the normative behavior for that situation, aid in his interpretation of the stimulus person's behavior. Yet with so many different possible sources of information available, it is difficult to determine what each source may contribute to any judgment, unless some are held constant or eliminated. In evaluating nonverbal behavior as a source of information we have eliminated verbal and vocal cues, and have held constant the observers' knowledge of the situation as given in the single set of instructions. Yet having given observers this information in the instructions about the situation makes it impossible to specify exactly what type of information they were utilizing in reaching their stress-catharsis judgment.

Judges had been specifically instructed to guess when the pictures had been taken. The terms *stress* and *final* were not defined as emotional experiences of the subject, per se, but were explained by a description of the interviewer's behavior in the two phases and how this may have affected the subject. Obviously the nonverbal cues shown in the pictures did not directly transmit the information *stress* or *final*, but this response of the judge represented an end point in a decisional chain in which he evaluated the cues in the photographs and then translated this evaluation

into the terms dictated by the experiment: *stress* or *final*. A judge's basis for this evaluation remains unknown, although at least three possibilities will be mentioned.

A judge knew what the interviewer was trying to do in each phase, and may have evaluated each photograph by thinking, "Does this look as if the interviewer has just attacked the subject's motivation for choosing his vocation, or as if the interviewer has just apologized for his behavior during stress?" The judge could also have made some inferences about the subject's affect, and approached each photograph with the question, "Does this look as if the subject is angry, hurt or apprehensive, or does it look as if the subject is relieved, relaxed and happy?" Another basis for evaluation would entail inferences about the relationship between interviewer and subject during the two phases; a judge might have thought, "Does this look as if they are distant or close; is the subject alert and paying attention, or withdrawn and uninterested?" Thus, three possible bases of evaluation may be distinguished: knowledge of the interviewer's behavior given in the instructions, inferences about the subject's affect, and inferences about the relationship between interviewer and subject. These three bases are not independent, but may have been used jointly by a judge, may each have been used by different judges, or may each have been used by the same judge when viewing different photographs.

The fact that the instructions gave the purpose of the interview, and an account of some of the interviewer's behavior and some of the subject's reactions, makes it impossible to specify which of these classes of information were linked to the nonverbal cues in identifying the stress and catharsis pictures. The experiment now to be introduced, however, eliminated any reference to any aspect of the interview and avoided mention even of the fact that an interview had been conducted, in an attempt to focus upon whether nonverbal behavior provides one specific class of information, affect.

Experiment XI

Problem. The major purpose of this experiment was to determine whether affect ratings of the photographs would differ between stress and catharsis pictures. By avoiding mention of the fact that the photographs were taken during an interview, Experiment XI tested whether nonverbal behavior alone, free of any aids in interpretation from knowledge of situational context, provides information about a person's affect which reflects the experiences we attempted to induce during the stress interviews. Thus this experiment parallels many of the past studies in the literature of the judgment of emotion. But, in place of the criterion of whether judge ratings match the intent of an actor to pose an emotion, our criterion was whether judge ratings of emotion differed for the two different interview phases.

A second purpose of this experiment was to verify the distinction

drawn previously, between consistently accurate, ambiguous, and consistently inaccurate photographs. Where photographs were randomly selected from the stress and catharsis phases of the interview, we could expect all three types of photographs to be represented. A difference in the affect ratings between stress and catharsis photographs randomly selected could be anticipated, but because of the presence of ambiguous and inaccurate photographs the difference should not be large. Where, however, only consistently accurate photographs were selected, then the difference in the affect ratings should be much larger.

Hypotheses: (1) Judge ratings of affect shown in the photographs will differ for the stress and catharsis photographs. (2) There will be a larger difference in affect ratings between stress and catharsis photographs selected from the consistently accurate category than from a sample of the two phases randomly selected from the entire range of photographs.

Method. This experiment is one of a series of studies (Ekman, 1964b), not to be presented here, utilizing Schlosberg's three dimensions of emotional expression to study the differences in communication between head and body cues.

Two samples of photographs were selected for ratings: a representative and a highly communicative sample. The representative sample was obtained by selecting 12 photographs at random from each of the five stress interviews: six of the pictures were taken from the stress phase, six from the catharsis phase. The 60 pictures from the five interviews were arranged into a randomly determined sequence. The highly communicative sample was composed of photographs for which more than 80% of the judges in the first nine experiments had correctly identified the interview phase; if more than one group of judges had seen a picture, then only pictures which met this criterion for each group of judges were chosen. Six pictures each from A, B, and C met this criterion; the twelve highest pictures within this criterion were chosen from D, and E; in each case half were stress and half catharsis photographs. Thus, the highly communicative sample was composed of 42 consistently accurate pictures; they were arranged in a randomly determined sequence for presentation to judges.

Judges rated each photograph on three nine-point scales: from pleasant to unpleasant, from tension to sleep, and from attention to rejection. Schlosberg's definitions of the three dimensions were used *verbatim* as follows:

> ***Pleasantness-Unpleasantness.*** You are to rate each picture on a 9 point scale where 1 indicates that the person in the photograph is feeling about as UNPLEASANT or unhappy as imaginable, and 9 indicates the maximum PLEASANTNESS.
>
> ***Attention-Rejection.*** You are to rate each picture on a 9 point scale where 9 indicates that the person in the photograph is feeling the maximum ATTENTION, as if the person is making every effort to see something. A rating of 1 indicates that the person is feeling the maximum REJECTION,

as if the person is trying to shut out or keep out any stimulation. Inattention is not the true opposite of attention, but occupies a position of about 5, midway between ATTENTION and REJECTION.

Sleep-Tension. You are to rate each picture on a 9 point scale where 1 indicates that the person in the photograph is feeling the complete relaxation of SLEEP, whereas 1 would be given to the most 'emotional' expression you can imagine, in which the person is very excited and shows maximum TENSION.

Schlosberg has suggested that these three scales represent the basic dimensions of emotional expression; his definitions of the scales and other aspects of his procedure were followed. Judges saw the photographs three times, rating each photograph on one of the scales during each presentation. The order in which the dimensions were considered by the judges was balanced. The individual photographs were shown for twenty seconds each during each presentation.

There was a slight overlap between these two photographs samples. About 10% of the representative sample was also included in the highly communicative sample.

Subjects. Twenty-six college freshmen served as judges of the representative sample of 60 photographs; seventeen college freshmen judged the communicative sample of 42 photographs.

Results. The judgments were analyzed by first obtaining a median rating for each judge across all of the stress photographs from the five interviews, the same for the catharsis photographs, and the difference between the two medians. A Wilcoxon matched-pairs signed-ranks test was then performed on the difference scores for all judges. This analysis was performed separately for the judges who viewed the representative sample and those who viewed the highly communicative sample. Table 10 summarizes these data, showing the median stress score across all stress photographs and across all judges, a similarly derived score for catharsis, and the median difference between the two. For both representative and communicative samples there was a significant difference on Pleasantness-Unpleasantness dimension ratings and on Attention-Rejection ratings, and no difference on Sleep-Tension ratings.

It is reasonable that *stress* was rated as more unpleasant than was *catharsis,* and as more "rejecting" than "attentive." However, the Attention-Rejection findings are questionable in that a high intercorrelation was found between Attention-Rejection ratings and Pleasantness-Unpleasantness ratings, and the difference obtained was small. The lack of difference between stress and catharsis photographs on Sleep-Tension ratings we have interpreted as signifying that the intensity of the affect states experienced during stress and catharsis was not different. Thus, the results shown in Table 10 support the first hypothesis on the Pleasantness-Unpleasantness dimension that there would be a difference in affect ratings for the stress and catharsis photographs, and the second hypothesis, that this difference would be larger for the highly communicative sample than for the representative sample.

Table 10 Median ratings of stress and catharsis in a representative and a communicative sample of photos in Experiment XI

	Pleasant (9) - Unpleasant (1)		Attention (9) - Rejection (1)		Sleep (1) - Tension (9)	
Sample:	Repr.	Comm.	Repr.	Comm.	Repr.	Comm.
Stress photos	4.33	3.89	5.93	5.75	5.60	4.75
Catharsis photos	5.68	6.89	6.27	6.43	5.61	4.88
Difference (catharsis-stress)	+1.25***	+3.00***	+.34***	+.68*	+.01	+.13

*$p < .05$
***$p < .001$

The data were further analyzed in order to examine the responses to each individual photograph from each interview. The median and interquartile range on the pleasantness scale was calculated for each individual photograph across all judges who had viewed that picture. Figure 4 shows these data for the representative sample, Figure 5 for the highly communicative sample. The difference predicted in the second hypothesis between the ratings on the representative and the highly communicative sample can be easily seen by comparing these two figures.

Discussion. This experiment differed from the others in that judges were deprived of any knowledge of the context within which the nonverbal behavior occurred, and were required to judge the affect of the stimulus person rather than guess when the pictures were taken. The fact that judges ascribed different affect to the pictures from the two

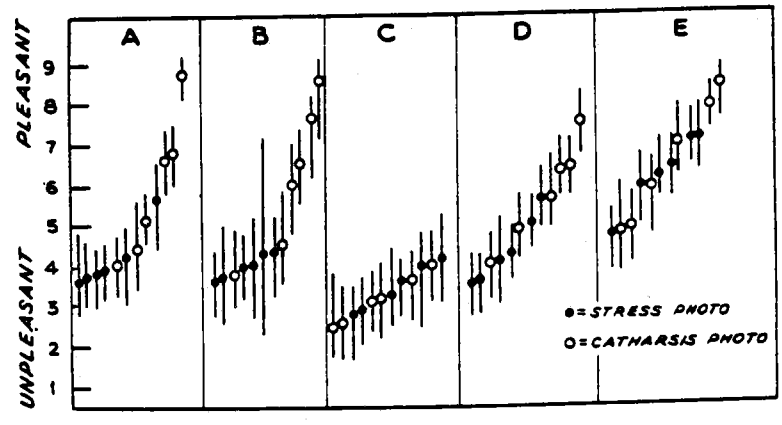


Figure 4. Median and interquartile range on pleasantness for each photograph in each interview for the random sample.

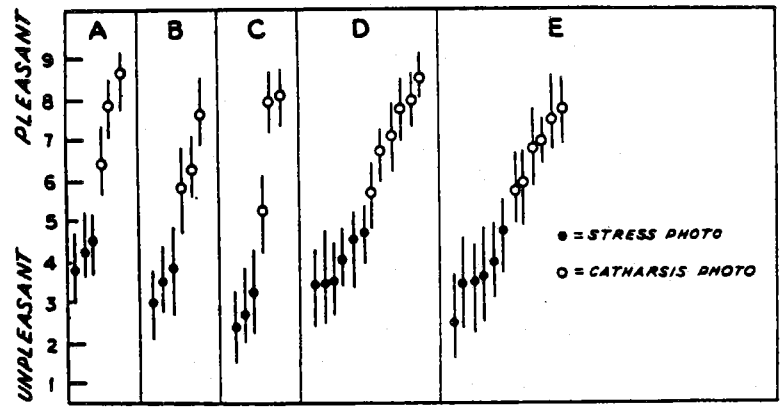


Figure 5. Median and interquartile range on pleasantness for each photograph in each interview for the highly communicative sample.

interview phases without any knowledge of the interview strengthens our conclusion that nonverbal behavior communicates accurate information. Experiment XI has shown that judges are able to interpret affect accurately from photographs without any knowledge of the situation, with the criterion of accuracy specifying that judge ratings of affect differ in an expected fashion for the two phases of the interview. It may be that other types of information, in addition to affectual, are communicated by nonverbal behavior, and that these other types were also utilized by the judges. A later discussion will introduce a speculative formulation about the range of classes of information communicable by nonverbal behavior.

Experiment XI served also to further verify that there were differences between the consistently accurate, ambiguous, and consistently inaccurate photograph categories. As had been predicted, there was a much greater difference in the affect ratings of the highly communicative sample (the consistently accurate photographs) than of the representative sample (containing inaccurate and ambiguous pictures as well as accurate ones).

A further verification of the distinction drawn between consistently accurate, ambiguous, and consistently inaccurate photographs was made by examining the Pleasantness-Unpleasantness ratings and the percentage of correct identification for the stress/catharsis task for each of the photographs in Experiment XI. There was an almost perfect linear relationship, such that with the stress photographs the ratings moved from Pleasantness to Unpleasantness as the percentage of correct identification varied from 0 to 100%. In like fashion, the ratings of the catharsis photographs moved from Unpleasantness to Pleasantness as the percentage of correct identification varied from 0 to 100%.

Figures 4 and 5, showing the results for each individual photograph

from each interview, draw our attention again to the differences found for the different stimulus persons in the five interviews. These data will be discussed next.

C. INDIVIDUAL DIFFERENCES IN NONVERBAL BEHAVIOR

The differences in results for the five stimulus persons will now be considered, to establish that the differences obtained can be attributed to actual differences between individuals in the clarity of their nonverbal communicativeness.

In the discussion of Tables 2 and 3 it was noted that low levels of accuracy were achieved for persons B and C, in contrast to A, D, and E. These differences were statistically significant.

In the discussion of Experiments I-IV, differences in the actual interview experience were raised as a possible explanation of such results. The interview experience might have been different for at least two reasons: either the interviewer's tactics varied, or the stimulus persons differed in their method of handling him. If the interviewer was too harsh during stress, and lost his knack of bringing about catharsis, then the subject's experience would have been uniformly stressful and judges would not have been able to differentiate the photographs even if the nonverbal behavior clearly carried the message about the experience. Similarly, if the interviewer could not bring himself to attack, then a subject's experience would have been rather non-stressful throughout, and again, even if the nonverbal behavior clearly reflected the subject's experience, judges would not have been able to differentiate the pictures from the two phases. Even if we assume that the interviewer performed similarly in all five interviews, the subject's method of handling the experience might have differed. If, for example, a subject was very well defended, accustomed to such assaults and skilled in parrying the interviewer, then his experience during stress might not have been very unpleasant, and again judges would have had difficulty identifying stress and catharsis pictures. Finally, if a stimulus person lacked resilience, or was vulnerable to attack, then the whole experience might have been miserable, and again judges would not have accurately identified the photographs.

In order to interpret the results on the five stimulus persons as due to some variable associated with clarity of nonverbal communication, we must find a measure which will reflect only differences in nonverbal sending, not these possible differences in the interview experience. One solution is to analyze the consistency versus ambiguity of information communicated, rather than accuracy. The consistency measure, as indicated earlier, is not contaminated by differences in the interview experience, since inaccurate photographs (i.e., stress photographs called catharsis and vice versa) are also considered consistent. The consistency measure, it may be recalled, was based on judge responses to each photograph. Pictures which elicited between 40 and 60% correct identifications were labeled ambiguous; all other pictures were considered

consistent, accurate or inaccurate. In Table 7 the proportions of consistent and ambiguous photographs for the five interviews were shown.

A uniformly stressful or uniformly cathartic interview experience would be reflected in a relatively high consistently inaccurate score, but would not contaminate the ambiguous category. Table 7 shows that Interview C had no more ambiguous photographs than the other interviews; instead the low level of accuracy achieved in the experiments seems to be due to the large number of consistently inaccurate photographs. In Interview B, on the other hand, not only does the stimulus person seem to have had a different interview experience (as shown in the proportion of inaccurate photographs), but more of his photographs were ambiguous than any of the others.

The greater ambiguity of stimulus person B's nonverbal behavior can also be inferred from the data in Experiment XI, shown in Figures 4 and 5. Both the median and the interquartile ranges for each of twelve photographs are shown on the Pleasantness-Unpleasantness scale for each of the five interviews. The medians reflect possible differences in individual experience within the interviews, and will be discussed below. The interquartile ranges, however, tell us how great was the spread of interpretation by judges for each of the pictures seen. We should expect that the interquartile ranges on B would be higher than the others, since we have noted that on the stress/catharsis task his behavior was more ambiguous. Inspection of Figures 4 and 5 shows that this is the case.

It is possible that the absolute median value might limit the size of the interquartile ranges, and therefore the best comparison of the interviews would hold the median value constant. Table 11 shows the average interquartile range for each interview, when only photographs with a median between three and five are considered from all interviews. This range of score values was chosen because it approaches the middle of the Pleasantness-Unpleasantness scale, and because it maximized the number of photographs which could be considered from each interview. As expected, B's photographs resulted in the largest average interquartile range.

The median values on the five interviews can serve to clarify some of our earlier impressions about the differences in interview experience among the five stimulus persons. Person B, whose behavior we have found to be most ambiguous, still received a number of different Pleasantness-Unpleasantness ratings, and most of his stress photographs were rated as

Table 11 Average inter-quartile range of photos between 3 and 5 on pleasantness-unpleasantness

	A	B	C	D	E
Number of photos	7	8	8	6	3
Average inter-quartile range	1.79	2.38	1.83	1.65	1.82

more *unpleasant* than his catharsis pictures. Person C, however, was rated as feeling uniformly *unpleasant;* there was no difference in the ratings of his stress and catharsis pictures. Finally we might note that person E was generally more highly rated as feeling *pleasant* than anyone else; even many of the photographs from the stress phase were rated as *pleasant.* What we are now considering is not how ambiguous the behavior was, not the spread of interpretations of any given photograph, but instead the range of different affects observed. While persons C and E are rated quite differently from the others, one as being typically *unpleasant* and the other as *pleasant,* we can not determine whether this represents a baseline difference in affect between the two, or in their method of handling the experience, or in the interviewer's behavior.

There is no theoretical import in having shown that person B's nonverbal behavior is ambiguous; the stimulus persons were not selected on any personality variable, and five persons is too small a sample for drawing inferences about personality variables associated with nonverbal communicativeness. We have, however, shown two methods for evaluating individual differences in nonverbal communication: consistency/ambiguity of information communicated with a dichotomous judgment, and average interquartile range on a scale judgment of affect. These measures could be applied to the nonverbal behavior of individuals selected according to a personality variable.

CONCLUSIONS

Our question for study was defined in the introduction as whether nonverbal behavior communicates accurate information about the quality of an interpersonal relationship; and, more specifically, as whether a change in the overall relationship between two interactants would be accurately communicated through nonverbal behavior. The experiments reported and discussed have provided the basis for a positive answer. This finding allows us to infer two propositions: (1) that nonverbal behavior systematically changes as a function of a gross modification in the quality of the relationship between two people; and (2) that such nonverbal behavior is communicative and can in part be accurately understood. The judges' accuracy on the first nine experiments would not have been possible unless both of these propositions are true. If there had been no systematic changes in the nonverbal behavior of the five stimulus persons, then judges would have performed no better than chance. The fact that these were unsophisticated receivers, with no prior experience in observing stress interviews, shows that the behavior which occurred in these standardized interviews is not unique to the laboratory or to these particular stimulus persons. Accuracy required that the particular body positions and facial expressions shown in the photographs be familiar to the receiver, and similarly interpreted by a number of such observers.

With reference to the literature reviewed at the outset of this paper, our findings are in general agreement with the findings of Dittmann, Exline,

Sainesbury, Rosenthal, Maccoby, Jecker et al., and Mahl, all of whom have found that spontaneous interactive nonverbal behavior either indicates or communicates information about a class of events, such as mood, diagnostic features, understanding, verbal content, experimenter bias, psychodynamics. Our research has added the evidence that nonverbal behavior can accurately communicate information relevant to a gross change in the relationship between the interactants. In light of the earlier research on expressive behavior, which suggested that accurate information depended in large part on some knowledge of the situational context within which the behavior was emitted (e.g., Munn, 1940), the results of Experiment XI are worthy of particular note. Accurate communication through nonverbal behavior, measured by predicted differences in affect ratings of behavior in the two parts of the interview, is possible without any knowledge about any feature of the situational context.

Most of the samples of nonverbal behavior were found to communicate consistent information to receivers. This would not be entirely surprising if the behavior had been posed, or the sender had known that he should so act that at some later time others could interpret his behavior when limited to the nonverbal channel alone; but this was not the case. Instead our statement that most of the nonverbal behavior emitted is not ambiguous, but rather carries consistent information, pertains to the more usual circumstances of interpersonal behavior—when individuals are behaving spontaneously and concentrating in large part on the verbal dialogue. The ratio of consistent to ambiguous information communicated to a group of receivers was noted as one method of discriminating between senders which is relatively free of many of the artifacts usually associated with reliance solely on a measure of consistently accurate information. A second method of measuring individual differences in clarity of nonverbal communicativeness was described and applied to the data. This measure, the average interquartile range on scales of affect, utilizes a different response system which permits an assessment of the spread in receiver interpretations of a given sender's nonverbal behavior, again relatively free of some of the artifacts more usually found in reliance upon accuracy scores.

In considering what specific information may be communicated by the photographs which would allow accurate identification of the interview phase, a number of possibilities were outlined, and evidence was obtained that at least one of the classes of information, impressions about affect, is communicated by spontaneous nonverbal behavior. Importantly, these affect ratings relate to the circumstances of the interpersonal relationship, even though those circumstances are not known by the persons who make the affect judgments. But, there are many other routes, many other possible classes of information, which may have been communicated by the nonverbal behavior to permit judges' accurate determinations. This chapter will close with a speculative formulation of the classes of information which may be communicated by nonverbal behavior.

A FORMULATION OF CLASSES OF INFORMATION COMMUNICATED BY NONVERBAL BEHAVIOR

In most situations nonverbal behavior is seen in the context of some knowledge of the situation and awareness of the concomitant verbal behavior. Though sometimes the verbal behavior is not heard by the observer, some features of the situation in which the behavior was emitted are usually known to him. Very rarely in real life is nonverbal behavior observed without any knowledge of the situation; usually in seeing another person's nonverbal cues we also learn something about his situation, and it is only in experiments that an observer is given the opportunity to judge nonverbal cues without having any other knowledge. Our discussion will consider: (A) what the observer can learn when he is completely deprived of any cues other than nonverbal ones, in order to establish a base line about what may be contributed by this source alone; (B) how these classes of information are interpreted when the observer does know something about the situation; and (C) how nonverbal behavior can function in relation to a verbal message.

Before proceeding, it might be wise to restate our use of the term *communication.* The communicative value of a nonverbal act is established by determining that a group of receivers will similarly interpret the sender's behavior. To say that the sender's nonverbal act is then communicative does not imply that the sender intended to communicate, nor does it imply that the communication is in any way accurate, for the receivers may be foolish, prejudiced or for some other reason may have completely misunderstood the significance of the act; but the act nevertheless communicates, since they, the receivers, show agreement in their interpretation of it. In no sense do we assume that there is any equivalence between the nonverbal cue and that which it communicates; while scratching the head may be found to communicate nervousness or anxiety or contemplation, there is no reason to suppose that head scratching is necessarily equivalent to, an attribute of, or an expression of any of these states. So cautioned, let us proceed with our discussion.

A. CLASSES OF INFORMATION FROM NONVERBAL BEHAVIOR ALONE

Affect: Experiment XI showed that nonverbal behavior alone can provide accurate information about affect. One interpretation of the results from Experiment X would be that nonverbal behavior may sensitively track moment-to-moment changes in affect.

Verbal-symbolic: Nonverbal behavior may provide publicly understood symbolic information through what are usually called gestures. The distinctive feature of a gesture is that the nonverbal act usually means very little in and of itself, but arbitrarily or by analogy has been assigned a precise symbolic meaning. It is because of this clearly accepted verbal

translation that we have labeled this class of information as verbal-symbolic.

Psychodynamic and diagnostic clues: The reports of gifted clinicians suggest that nonverbal behavior may provide symbolic information which is private, rather than public, in that it is recognized and interpreted only by rather special observers. Mahl (1959), for example, was able to infer marital problems from a patient's play with her wedding ring, or problems in managing aggressive impulses from hand clasping positions.

Instrumental acts: Nonverbal behavior is directly involved in the pursuit of certain instrumental activities even when an individual sits in a chair, and provides information that the person is tying his shoe, or scratching his head, or smoking.

Portrayals, dramatizations, reenactments: Nonverbal behavior can be used to act out in miniature or in detail a past, present, or future event.

Demography: Visual appearance, although at times deceptive, rather clearly states the age, sex, and perhaps through dress the social status of an individual. While not precisely demographic, related information about intelligence, aesthetic quality, values and occupation can be inferred from nonverbal behavior.

Style: The rate, rhythm and type of bodily activity can lead to inferences about expressiveness and temperament.

This list is not meant to be exhaustive, but to suggest some different kinds of information which may be communicated by nonverbal behavior. Obviously no claim is made that nonverbal behavior provides accurate information about each class of information, although it is probable that through the operation of stereotypes a reasonable amount of *consistent* information is communicated about each.

In some cases we have linked specific forms of nonverbal behavior to a particular category of information, and it is conceivable that this might be more systematically explicated. The problem is a difficult one, especially since a given cue can provide more than one type of information; e.g., an instrumental act may tell us not only what a person is doing at the moment, but also about his affect and perhaps about his psychodynamics.

B. ADDITIONAL INFORMATION AVAILABLE WHEN NONVERBAL BEHAVIOR IS INTERPRETED WITHIN KNOWLEDGE OF THE SITUATIONAL CONTEXT

If the observer also knows something about the situation in which the behavior occurred, then more specific inferences can be drawn within each of the classes of information. Usually the observer has such knowledge available, even if he cannot hear the verbal communication. There are many situations in which there is a shared intent to communicate, but the situation prevents verbal exchange: hitch-hiker and motorist, pilot and landing crew, charade player and audience, hunters stalking prey, etc.

If extensive communication is necessary, then gestures, improvised forms of verbal-symbolic behavior, portrayals, and instrumental acts will predominate. There are other situations in which the sender is not talking, does not intend to communicate to the observer, or is involved in an exchange with another person, where the observer can see but not hear. An example is the hidden observer; or the visual eavesdropper watching a conversation across the room at a cocktail party; or the therapist trying to understand the mute or resistant patient. In such situations the observer will attempt to utilize all of the classes of information available, although such typically intention-based categories as verbal-symbolic and portrayals will not be present.

We will consider only interpersonal situations, exploring first how the observer's interpretations of nonverbal behavior may be aided by knowledge of an abstract aspect of an interpersonal relationship: whether it is representative or atypical of the stimulus person's relationships.

If the nonverbal cues provide information about affect, and we also know that the situation is one typical of a person's interpersonal relationships, then it is possible to draw further inferences about his general level of satisfaction, and his usual or enduring moods. If, for example, we note that he looks sad, and we also believe that we have seen a representative sample of his usual affective reactions in his interactions, inferences could be drawn about the presence of a depressive disorder. Probably closely related to affect is information about the *quality of the interpersonal relationship*—whether it is close or distant, whether the person is involved and interested or detached and withdrawn. Knowledge about these two classes of information, affect and relationship quality, can lead to further inferences about the person's typical *role*—whether he is compliant or assertive, pliable or rigid, open or defensive, dominant or submissive, etc. If the observer knows that the sample of behavior shown is typical of the stimulus person's relationships, then such information about relationship quality and role can lead to further judgments about the general tenor of his interpersonal relationships, and related personality formulations based on assumptions about these relationships.

If the role of the other participant is known, whether that of spouse, employer, stress interviewer, etc., then many more specific inferences can be drawn from the classes of information provided by nonverbal cues. Information about affect can tell how satisfied the person is in his marital adjustment or work situation, what moods characterize his friendships, etc. Similarly, information about relationship quality can suggest that he is distant with employers but close with spouse, or any other such possibility. And finally, information about interpersonal role could suggest subservience with employer but dominance with wife as the typical roles.

To summarize, knowledge of the situation within which nonverbal behavior is emitted can greatly expand the interpretations of nonverbal cues. If it is known that the nonverbal behavior occurred during an

interpersonal relationship, two new classes of information can be inferred from nonverbal behavior—*relationship quality* and *role*. If the observer knows that the sample of nonverbal behavior is representative of the stimulus person's usual relationships, then information about affect, relationship quality, and role can lead to more specific inferences about adjustment in different types of interactions, and formulations about the general style of interpersonal relationships and associated psychodynamic and diagnostic features.

C. HOW NONVERBAL BEHAVIOR MAY PROVIDE INFORMATION RELATED TO VERBAL BEHAVIOR

In other research, also based on the stress interviews (Ekman, 1964a), we have shown that a very specific moment-to-moment relationship between verbal and nonverbal cues can be accurately recognized by an observer. The relationship between these two channels of communication is complex; nonverbal behavior can serve a variety of communicative functions in relation to verbal behavior. Seven functions will be considered here: (1) repeating, (2) contradicting, (3) substituting for a verbal message; (4) reflecting the person's feeling about his verbal statement; (5) reflecting changes in the relationship; (6) accenting parts of the verbal message; and (7) maintaining the communicative flow.

Nonverbal behavior can simply *repeat* the substance of a verbal message. If the verbal behavior describes an affective reaction, the nonverbal behavior can repeat the affect. If the verbal behavior describes a certain event or course of action, the nonverbal behavior can be an action portrayal of the event or action. Gestures can also repeat some aspect of the verbal message. These nonverbal repetitions of verbal messages can serve to emphasize the message; their exaggeration or understatement or lack of appearance can lead to inferences about style, e.g., expressiveness, warmth.

Nonverbal behavior can directly *contradict* the content of the verbal message. The most obvious case is an affect shown nonverbally which directly contradicts the verbal message. Other information carried by nonverbal behavior can also contradict the verbal message. The person who says "yes" and shakes his head "no" provides an instance of a gesture contradicting a verbal message. The individual who verbally states his control over a situation and nonverbally drops what he is holding, trips, etc., provides an instance of instrumental actions contradicting a verbal message. Such nonverbal contradictions of verbal behavior can lead to further inferences about areas of conflict and attempts to inhibit or control communication.

Nonverbal behavior can also be a *substitute* for a specific word or phrase in a verbal message. A nonverbal expression of affect can replace the verbal statement and be directly embedded in a verbal message describing how the person feels. Similarly, instrumental acts, gestures, portrayals, can all be used as substitutes for part of a verbal message.

Nonverbal communication can indicate a person's *feeling about his verbal statement;* e.g., shyness, embarrassment, pride, can all serve to qualify what is being stated verbally. In a similar way, nonverbal cues pertaining to the relationship quality can be informative of how *changes in the relationship* are affecting or being affected by the verbal level of discourse. A change in posture in a chair from a relaxed to a more formal or stiff position, while a patient verbally states a highly charged theme, might be communicating that the relationship is bordering on more tender areas and the sender wishes to retreat to a more stereotypical superficial relationship. Certain actions, usually of the head or hands, can be used rhythmically to *accent* or underline certain words. Certain rather minimal nonverbal cues, head nods, eye movements, shifting of position, can serve to *maintain the communicative flow.* These nonverbal cues serve to signal when the speaker needs feedback from the receiver, when he is nearly finished speaking and will allow the receiver a chance to communicate verbally; from the receiver they can indicate that he agrees and the sender need not stop his line of discourse, that he cannot wait much longer to get in his own ideas, etc.

To summarize, seven classes of information communicated by nonverbal behavior alone are presented: affect; verbal-symbolic; psychodynamic and diagnostic clues; instrumental acts; portrayals; demography; style. If the observer also knows that the behavior was emitted during an interpersonal relationship, then nonverbal cues may also communicate information about the relationship quality and the stimulus person's role. Further knowledge of how representative the behavior is, and the role of the other participant in the relationship, can lead to interpreting from nonverbal cues information about the nature and style of interpersonal relationships. Finally, most of the classes of information provided by nonverbal behavior can serve to repeat, contradict, or substitute for a verbal message, as well as accent certain words, maintain the communicative flow, reflect changes in the relationship in association with particular verbal messages and indicate a person's feeling about his verbal statement.

REFERENCES

Allport, G. *Pattern and growth in personality.* Chapter 19. New York: Wolf-Rinehart, 1961.

Bateson, G. Personal communication. 1962.

Birdwhistell, R. L. *Introduction to kinesics.* Louisville: University of Louisville Press, 1952.

Brengelmann, J. C. Expressive movements and abnormal behavior. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1963.

Bruner, J. S, & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge: Addison-Wesley, 1954.

Buehler, R. E., & Richmond, Jo F. Interpersonal communication behavior analysis: a research method. *J. Communic.,* 1963, *13,* 3, 146-155.

Davitz, J. R. (Ed.) *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Dierssen, G., Lorenc, M., & Spitalerl, R. M. A new method for graphic study of human movements. *Neurology*, 1961, *11*, 610-618.

Dittmann, A. T. The relationship between body movements and moods in interviews. *J. consult. Psychol.*, 1962, *26*, 5, 480.

Dittmann, A. T. Personal communication. 1963.

Ekman, P. A methodological discussion of nonverbal behavior. *J. Psychol.*, 1957, *43*, 141-149.

Ekman, P. Body position, facial expression and verbal behavior during interviews. *J. abnorm. soc. Psychol.*, 1964a, *68*, 3, 295-301.

Ekman, P. A comparison of the information communicated by head and body cues. Paper read at American Psychological Association Convention, 1964b.

Exline, R. V. Explorations in process of person perception: visual interaction in relationship to competition, sex, and need for affiliation. *J. Pers.*, 1963, *31*, 1.

Giedt, F. H. Comparison of visual, content and auditory cues in interviewing. *J. consult. Psychol.*, 1955, *19*, 6, 407-416.

Giedt, F. H. Cues associated with accurate and inaccurate interview impressions. *Psychiatry*, 1958, *21*, 4, 405-409.

Hoffman, M. The effects of training on the judgment of non-verbal behavior. An experimental study. Doctoral dissertation. Harvard University, Cambridge, Mass., 1964.

Jecker, J., Maccoby, N., Breitrose, H., & Rose, E. Teacher accuracy in assessing cognitive visual feedback from students. *J. appl. Psychol.*, 1964, in press.

Jones, F. P., & Narva, M. Interrupted light photography to record the effect of changes in the poise of the head upon patterns of movement and posture in man. *J. Psychol.*, 1955, *40*, 125-131.

Jones, F. P., O'Connell, D. N., & Hanson, J. A. Color coated multiple image photography for studying related rates of movement. *J. Psychol.*, 1958, *45*, 247-251.

Jones, F. P., & Hanson, J. A. Time-space pattern in a gross body movement. *Percept. mot. Skills*, 1961, *12*, 35-41.

Klein, Z. E. The nonverbal communication of feelings: a review of the literature. Unpublished manuscript, 1963.

Maccoby, N., Jecker, J., Breitrose, H., & Rose, E. Sound film recordings in improving classroom communications; experimental studies in nonverbal communication. Report, Institute for Communication Research, Stanford University, 1964.

Mahl, G., Danet, B., & Norton, Nea. Reflection of major personality characteristics in gestures and body movements. *Amer. Psychologist*, 1959, *7*, 357.

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

Rosenthal, R. Experimenter attributes as determinants of subjects' responses. *J. proj. Tech.*, 1963(a), *27*, 3, 324-331.

Rosenthal, R. On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. *Amer. Scientist*, June 1963 (b), *51*, 2, 268-283.

Sainesbury, P. A method of recording spontaneous movements by time-sampling motion pictures. *J. Ment. Sci.*, 1954, *100*, 742-748.

Sainesbury, P. Gestural movement during psychiatric interview. *Psychosom. Med.*, 1955, *17*, 458-469.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Sullivan, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.