

Relative Importance of Face, Body, and Speech in Judgments of Personality and Affect

Paul Ekman and Wallace V. Friesen
University of California, San Francisco

Maureen O'Sullivan
University of San Francisco

Klaus Scherer
University of Giessen
Giessen, West Germany

Three experiments correlated judgments made from observing single channels (face, body, or speech) with multiple channel judgments (face, body, and speech together; or face and speech together). Judges observed spontaneous behavior in two different types of interview situations. The single channel judgments that correlated most highly with the multiple channel judgments depended on the type of attribute being judged and the situation in which the behavior occurred.

Three different groups of determinants may be distinguished in making judgments about another person's emotional state, attitude, or personality. One group of determinants involves the observer who makes the judgment: how the observer's own emotional state, attitude, personality, skill, expectations, motivation, or involvement in the judgment task procedure might influence the observer's judgment of others. A second group of determinants involves the same considerations about the person being judged: how the stimulus person's own emotional state, attitude, and so on might determine the information provided to the observers. A third group of determinants is concerned with the sources of information made available to the observer: how different aspects of the stimulus person's behavior (e.g., voice quality, body movements, etc.) might influence judgments.

Many studies have focused on the third

group of determinants, examining whether nonverbal behavior contributes more than verbal behavior to an observer's judgments. In such research, observers have been exposed to different *channels*. Some have been allowed to hear and others to see a sample of behavior, or speech has been separated into content and voice quality channels and nonverbal behavior into face and body channels. It is difficult to reach conclusions from these experiments because of a number of methodological shortcomings.

Most studies presented only a few stimulus persons (usually only one or two) in an artificial interpersonal situation (an actor demonstrating an emotion). The question of whether the importance of a channel might depend on or interact with the type of information being judged, the characteristics of the observer making the judgment, or the characteristics of the stimulus person being rated has not been considered in most experiments. (Researchers who have studied some of these issues include Berman, Shulman, & Marwit, 1976; Bugental, Kaswan, & Love, 1970; Cline, Atzet, & Holmes, 1972; Ekman, 1965; Friedman, 1978; Scherer, Scherer, Hall, & Rosenthal, 1977; Vande Creek & Watkins, 1972.)

It is difficult to draw conclusions from this literature because so many different methods have been used to evaluate the importance

This research was supported by Grant MH 11976 from the National Institute of Mental Health. The preparation of this article was supported by a grant from the Harry F. Guggenheim Foundation and by Research Scientist Award MH 06092 from the National Institute of Mental Health.

Requests for reprints should be sent to Paul Ekman, Human Interaction Laboratory, Department of Psychiatry, University of California, San Francisco, 401 Parnassus Avenue, San Francisco, California 94143.

of a channel. The most popular methods have been to determine either which channel provides the most accurate judgments (e.g., Cline et al., 1972) or which channel provides the highest level of interrater agreement (e.g., Ekman, 1965). In this study, we chose a method that has been used less often: determining which channel correlates most highly with ratings made by observers who are exposed to the full audiovisual record.

This method most directly addresses the issue of how single channels contribute to the impressions formed when an observer is presented with the usual multichannel interpersonal input (simultaneous face, body, and speech behavior).¹ We did not expect one channel to contribute the most, as some previous investigators have reported (e.g., Mehrabian & Ferris, 1967). Instead, we hypothesized that the relative contribution of a channel would depend on other variables such as the attribute being judged or the type of social situation in which the rated behavior had occurred. Testing this hypothesis required that ratings be obtained on many different attributes by observers exposed to the behavior of a large sample of stimulus persons in at least two different interpersonal situations. Three separate experiments allowed replication across different groups of observers.

Experiment 1

Method

Stimulus materials. Black and white videotapes were obtained in a laboratory situation developed to study deception (Ekman & Friesen, 1974). Fifteen student nurses were videotaped in each of two standardized interviews. In both interviews, the subject watched a short film and answered an interviewer's questions concerning her feelings about it. In the *honest* interview, the subjects were in a relatively unstressful situation. Nature films designed to elicit pleasant feelings were shown, and subjects were instructed to describe their feelings frankly. In the *deception* interview, subjects saw a film showing amputations and burns, intended to elicit strong unpleasant affect. They were instructed to conceal negative feelings and to convince the interviewer they had seen another pleasant film. Thus, in the deception condition, the subjects viewed a highly stressful film and had the additional stress of deceiving the interviewer. A variety of data from other studies of another group of comparable sub-

jects in this situation (Ekman & Friesen, 1974; Ekman, Friesen, & Scherer, 1976) suggested that subjects are highly motivated to succeed in this deception and that the intended emotional states are elicited in each interview. These two interviews provided behavior samples from interpersonal situations that differed both in affect experienced and in frankness of expression.

Two-minute segments (almost the entire short interview) from each honest and each deceptive interview were edited in a randomized order onto two videotapes in such a way that each subject was shown in only one situation (honest or deceptive) on each videotape. Each tape showed 15 interviews, about half of which were honest and half deceptive. A given group of observers saw only one tape. The videotape showed a head-on view of the subject seated in a chair, with the entire face and body (including the feet) visible. Channel separation was achieved by blocking off the face on the video monitor for those shown only the body, blocking off the body for those shown only the face,² and turning on the audio and darkening the screen for those who heard only the speech. Speech thus included both what was said, speech content, and how it was said, voice quality. In the whole person condition, which was the criterion in this experiment, the entire monitor was exposed and the sound switched on.

Rating scales. Fourteen 7-point bipolar adjective scales were used (see Table 1). An attempt was made to include scales germane to each channel (e.g., awkward-natural for the body, likable-unlikable for the face, dominant-submissive for speech) as well as scales that might be relevant across channels. The scales have been grouped on the basis of factor analyses performed in a previous study (Ekman et al., 1976). Each of the first three groups of scales represents a stable factor, whereas the last four scales do not load consistently on any factor.

Observers and procedures. Eight groups of observers were used, two for each of the four sources (face, body, speech, whole person), so that no single observer would see a subject in both her honest and deceptive interview. These groups ranged in size from 11 to 22. Observers were recruited through a mailed invitation to participate in research on non-

¹ Subsequent reports will utilize other methods to study differences among channels: level of agreement reached by observers of a particular channel and the accuracy of their judgments.

² Obtaining a face and body channel separation by blocking one or another off the video monitor is convenient but runs the risk that the small size of the face so obtained might restrict the information ordinarily available to an observer. In another study (Ekman, Brattesanni, O'Sullivan, & Friesen, 1979), no difference was found between the judgments reported in this experiment on a small facial image and the judgments made by another group of observers who saw a face five times as large.

verbal communication that was sent to male and female adults who receive catalogs from the Extension Division of the University of California. The observers were not told that deception was involved in the interviews. The observers rated each stimulus person on all 14 scales immediately after each 2-min. segment.

Results

All analyses used the mean of each group of observers' ratings of each stimulus person on each scale. The data, then, consisted of 1,680 group means—4 means (face, body, speech, whole person) for each of 2 conditions (honest or deception) for each of 14 scales for each of 15 stimulus persons. Table 1 lists the Pearson *r*s between the mean ratings made by those observers exposed to the whole person condition and the mean ratings from the groups of observers of each of the separate channels. The correlations between the single channel and the whole person judgments were determined scale by scale over the 15 stimulus persons. The highest correlation between each channel and whole person judgment for each scale is italicized. Stepwise regression analyses using the mean ratings of the whole person as the criterion were also performed (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). These multiple correlations are also listed in Table 1. The columns in Table 1 give the results for Experiment 1 and also for Experiment 2. The rows give the findings for each scale for both experiments. The results in Table 1 are organized separately for the ratings of the behavior from the honest and deception situations.

Claims from the previous literature about the primacy of the visual over the verbal channel and the particular importance of the face as compared to body or speech (Argyle, Alkema, & Gilmour, 1971; Burns & Beier, 1973; DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978; Mehrabian & Ferris, 1967; Zaidel & Mehrabian, 1969) were not supported. No single channel (face, body, or speech) was consistently most highly correlated with the whole person judgments. When the behavior during the honest interview was rated, the degree of correlation between channel and whole person judgments varied with the type of characteristic (i.e.,

rating scale) being judged. The pattern of correlations will be discussed later, together with the results from Experiment 2.

Remarkably, when behavior during the deception situation was judged, judgments made from speech alone rather consistently had the highest correlation with whole person judgments. Judgments of speech in the honest situation had the highest correlation with the whole person judgment on 4 scales, and judgments of speech in the deceptive situation had the highest correlation with whole person on 12 scales. Chi-square analysis indicated that this difference was significantly greater than might be expected by chance, $\chi^2(1) = 9.33, p < .005$. Thus, the situational variable, whether the stimulus person's behavior had occurred during an honest or a deceptive interview, had a significant impact on which channel correlated most highly with the criterion.

Channels may correlate differentially with whole person judgments for many reasons. One reason is that some channels or some scales may be unreliable. This unreliability will restrict the degree of correlation possible with other variables. To determine the effect of the reliability of the judges' ratings on the results obtained, split-half reliabilities extended by the Spearman-Brown formula were obtained for each scale and each channel in both conditions. Reliabilities were comparable across all channels and therefore do not explain the results of Experiment 1. Median reliabilities ranged from .75 for the face in the deception condition to .88 for speech in the honest condition.

Before interpreting the findings from Experiment 1, it was necessary to rule out a simple, albeit theoretically uninteresting, explanation of why speech was so often most highly correlated with the criterion when behavior in the deception situation was judged. Although none of the observers had been told that the stimulus persons had been instructed to deceive, in the second minute of the deception interviews, the subjects had been asked if they were telling the truth. The observers of the face only and body only channels did not hear what was said, and deceit would not have been salient for them. Speech

and whole person observers would hear this question, and this shared exposure to doubts about veracity might account for the high correlations between their judgments. To eliminate this possibility, Experiment 2 included only the first minute of the interviews in which the questions asked were identical for both the honest and deception interviews and in which the issue of truthfulness was not explicitly raised by the interviewer.

Experiment 2

Method

The stimulus materials were the same as in Experiment 1, except that the interview segments were 1 min. rather than 2 min. in length. Eight new groups of observers were used. Students from psychology classes at the University of San Francisco volunteered to be observers in return for earning credit towards their grade. The groups varied in size from 10 to 17.

Results

The results from this experiment are listed in the columns headed Experiment 2 in Table 1. The overall pattern of correlations for the ratings of the honest interviews replicated the findings of Experiment 1. Again, no single channel was most highly correlated with the ratings of the whole person for all scales. The most important channel, the one most relevant to the impressions formed when people were subject to the usual interpersonal input (whole person ratings), varied with the type of information judged.

In the judgments of the honest situation, face, body, and speech all showed roughly equal, high levels of correlation with the criterion whole person judgments for the first cluster of scales (outgoing, expressive, sociable). Pooled correlations over all three scales for each channel were not significantly different from one another. Thus, for the honest situation, similar rankings about sociability or expressiveness can be derived from many channels. Each channel alone yielded about the same impression as simultaneous exposure to all channels.

This is in marked contrast to what was found for the second group of scales (calm, natural, stable, relaxed). In the judgment of

behavior in the honest situation, either face or body correlated highest with the criterion. Over the four scales in this group, pooled correlations between face and whole person were higher (Experiment 1, $p < .002$; Experiment 2, *ns*) than pooled correlations between speech and whole person judgments. The pooled correlations between body and whole person were also higher (Experiment 1, $p < .001$; Experiment 2, $p < .032$) than the pooled correlations between speech and whole person judgments. Thus, information concerning a person's calmness or relaxation derived from either the face or body would give about the same impression as exposure to all channels when the stimulus persons were in a situation in which they were honestly describing a positive experience.

Judgments on the third group of scales (honest, sincere, trustworthy) produced correlations with criteria that were generally lower than those for the other scales, and there were a number of inconsistencies in the relative size of the correlations found in Experiments 1 and 2. Finally, the judgments on the last four scales (dominance, likeability, felt pleasant, and acted pleasant) showed a pattern somewhat similar to that of the Group 1 scales. When behavior in the honest situation was observed, ratings of face, speech, and body were all equally highly correlated with the criterion.

The other major finding from the first experiment was essentially replicated. When the behavior during the deception situation was judged, judgments made by the observers exposed to speech most often had the highest correlation with the judgments made by observers of the whole person. For the honest situation behavior, speech judgments had the highest correlation with whole person judgments on 5 scales, and for the rating of behavior in the deception situation, speech judgments had the highest correlation on 10 scales. Although the chi-square, performed as in Experiment 1, did not quite reach the .05 level of significance, $\chi^2(1) = 3.59$, $p > .05$, the trend was in the same direction. As in Experiment 1, reliability of the ratings does not seem to be an important determinant of which channel correlates highest with whole person

ratings. Median reliabilities, determined as in Experiment 1, were not significantly different from one another and therefore do not explain the findings (median reliabilities ranged from .60 for the face in the honest condition to .86 for speech in the deception condition).

In Experiment 2, the mean age of the observers was some 20 years younger than the mean age of the observers in Experiment 1. Thus, the essential replication of findings in Experiment 2 shows that there is generality regardless of the age of the observers or whether the behavior sample is 1 or 2 min. in length.

Experiment 3

In this experiment, a face with speech (face-speech) condition was shown to observers and used as the criterion instead of whole person. This provided another opportunity to determine whether the channel

that correlated most highly with a criterion would depend on the scale being judged and whether the speech channel would be more important when behavior in the deception situation was judged.

Method

Two new groups of observers recruited from the same pool as Experiment 2 judged the videotapes used in Experiment 2 with the audio turned on and the body blocked off to provide the new criterion judgments of face-speech. In other respects, the procedure was the same as in Experiments 1 and 2. The face only and speech only ratings gathered in Experiment 2 were used for the single channel judgments to correlate with the new criterion ratings.

Results

Table 2 is organized like Table 1. The findings are consistent with the results shown in Table 1 for Experiments 1 and 2. Whether face or speech channel judgments were most highly correlated with the criterion (face plus

Table 2
Correlations and Multiple Correlations Between Single Channel Judgments and Face With Speech in Experiment 3

Judgment	Honest condition			Deception condition		
	Face	Speech	<i>R</i>	Face	Speech	<i>R</i>
Group 1						
Outgoing-inhibited	<i>.80</i>	.76	.911	<i>.88</i>	.85	.904
Expressive-unexpressive	<i>.73</i>	.71	.810	<i>.75</i>	.87	.897
Social-withdrawn	<i>.67</i>	.64	.746	<i>.77</i>	.76	.836
Group 2						
Calm-agitated	<i>.14</i>	.12	.210	<i>.31</i>	<i>.80</i>	.810
Natural-awkward	<i>.58</i>	.54	.676	<i>.65</i>	<i>.81</i>	.829
Stable-unstable	<i>-.07</i>	<i>.24</i>	.293	<i>.52</i>	<i>.63</i>	.686
Relaxed-tense	<i>.59</i>	.35	.610	<i>.43</i>	<i>.72</i>	.726
Group 3						
Honest-dishonest	<i>.60</i>	.15	.605	<i>.39</i>	<i>.60</i>	.723
Sincere-insincere	<i>.11</i>	<i>.23</i>	.233	<i>.52</i>	<i>.74</i>	.761
Trustworthy-untrustworthy	<i>.01</i>	<i>-.18</i>	.185	<i>.27</i>	<i>.29</i>	.357
Group 4						
Dominant-submissive	<i>.78</i>	<i>.81</i>	.892	<i>.68</i>	<i>.84</i>	.877
Likable-unlikable	<i>.26</i>	.11	.268	<i>.80</i>	.61	.809
Felt pleasant-felt unpleasant	<i>.17</i>	<i>.51</i>	.509	<i>.62</i>	<i>.79</i>	.790
Act pleasant-act unpleasant	<i>.68</i>	.57	.702	<i>.82</i>	<i>.77</i>	.862

Note. Decimal points are omitted. Italicized values are the highest correlations between single channel and criterion ratings.

speech) varied across the characteristics judged when the behavior from the honest situation was observed. For the behavior in the deceptive situation, speech channel judgments generally correlated most highly with face-speech judgments. In the honest situation judgments, speech ratings had the highest correlations with face-speech on 4 scales, and in the deception situation judgments, speech ratings had the highest correlation with face-speech on 10 scales. A chi-square performed as in Experiment 1, showed that this difference again was significant, $\chi^2(1) = 5.14, p < .025$.

Discussion

The results from all three experiments show that it would be unwise to claim that any one channel predominates in judging other people. Which channel predominates depends on what characteristic is being judged as well as on the interpersonal situation in which the judged behavior occurs. The claims in the literature that the face is most important or that the nonverbal-visual channel is more important than the verbal-auditory channel have not been supported. The question is more complex, at least if importance is understood in terms of the amount of correlation between ratings based on single channels and ratings of persons made when the usual verbal and nonverbal inputs are available.

It is not clear whether the difference between the results obtained in these experiments and those reported in the literature is due to (a) a design that evaluates importance of a channel by how well it correlates with impressions formed when people are exposed to the total array of cues available when judging others—a design not used in most other channel studies, (b) a larger array of persons judged and a larger array of scales than have been used in almost all previous studies, or (c) naturally occurring behavior rather than posed behavior, which has been used in most other channel studies.

Although all three experiments support the generalization that in judging behavior in the honest situation no single channel always cor-

relates most highly with the criterion, the replications are in terms of the observers who judged the behavior. Essentially the same set of interviews, all with females, was used throughout. Replication is needed with other stimulus persons and other social situations before it will be possible to know the generality of the specific pattern of correlations between single and multiple channels.

It is noteworthy that the pattern of correlations shifted so dramatically when the deceptive situation interview behavior was judged. It should be remembered that none of the observers were told that the subjects were instructed to deceive in some interviews and that in Experiments 2 and 3 the observers could not have heard any questions suggesting that honesty was an issue. We can offer two explanations as to why the impressions from speech correlated so highly with impressions formed from whole person or face-speech when behavior in the deception situation was observed.

Other studies (Ekman et al., 1976) of another group of nurses who were subjects in these honest and deceptive interviews found that voice pitch level went up in the deception situation. The heightened pitch and perhaps other vocal changes may have drawn attention to the auditory channel and away from the visual channel. Further research that separates the speech channel into speech content and voice quality (filtered speech) is needed to determine whether the observers of behavior in the deceptive situation are indeed relying more on the sound of the voice than on what is actually said. Evidence from other studies (Scherer, Koivumaki, & Rosenthal, 1972; Scherer et al., 1977) suggests that much information relevant to judgments of affect and personality can be made from noncontent aspects of speech.

The second alternative explanation is derived from Ekman and Friesen's (1969) theory of nonverbal leakage, which proposed that people learn to monitor and disguise the content of their speech more than their face, body, or voice quality. If that is so, and there is some indirect evidence for that presumption (Ekman & Friesen, 1974), the face, body, and voice quality might each contain contra-

dictory information. The content of speech, by contrast, should be more consistent (that is, well disguised), and perhaps the observers were most influenced by this consistency. Of course, there was no separate speech content condition in this experiment, and this interpretation would be correct only if the observers of the speech condition relied most on the speech content. A related idea is that if observers are confronted with contradictory information, they will rely most on the source that is most explicitly overlearned, namely, speech content.

The data from these experiments are relevant to another issue, quite apart from the importance of whether the behavior judged was from the honest or deceptive situation. The multiple correlations listed in Tables 1 and 2 suggest that there may be considerable information that is not derived from any one single channel but is the product of what is learned when observation cuts across channels. Apart from the first group of scales, where the multiple correlations were very high, there are a number of scales in which the combination of channels failed to account for half of the variance in the judgments based on impressions from the whole person or face plus speech.

This study has shown the complexity of the phenomena involved in judging another person's emotional state, attitudes, or personality. No simple statement can be made about the determinants of such judgments. The attribute being judged and the situation in which the behavior occurred markedly affected whether separate channels could account for the judgments made from the more usual interpersonal inputs (i.e., whole person or face-speech). No single channel was most important across attributes judged and situations observed. The differential weighting of nonverbal and verbal behavior was a function of the situation in which the behavior was shown and of the trait or characteristic being judged.

References

- Argyle, M., Alkema, F., & Gilmour, R. The communication of friendly and hostile attitudes by verbal and nonverbal signals. *European Journal of Social Psychology*, 1971, 1, 385-402.
- Berman, H. J., Shulman, A. D., & Marwit, S. J. Comparison of multidimensional decoding of affect from audio, video and audiovideo recordings. *Sociometry*, 1976, 39, 83-89.
- Bugental, D., Kaswan, J., & Love, L. Perception of contradictory meanings conveyed by verbal and nonverbal channels. *Journal of Personality and Social Psychology*, 1970, 16, 647-655.
- Burns, K. L., & Beier, E. G. Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of Communication*, 1973, 23, 118-130.
- Cline, V. B., Atzet, J., & Holmes, E. Assessing the validity of verbal and nonverbal cues in accurately judging others. *Comparative Group Studies*, 1972, 3, 383-394.
- DePaulo, B., Rosenthal, R., Eisenstat, R., Rogers, P., & Finkelstein, S. Decoding discrepant nonverbal cues. *Journal of Personality and Social Psychology*, 1978, 36, 313-323.
- Ekman, P. Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology*, 1965, 2, 725-735.
- Ekman, P., Brattesani, K., O'Sullivan, M., & Friesen, W. V. Does image size affect judgment of the face? *Journal of Nonverbal Behavior*, 1979, 4, 57-61.
- Ekman, P., & Friesen, W. V. Nonverbal leakage and clues to deception. *Psychiatry*, 1969, 32, 88-105.
- Ekman, P., & Friesen, W. V. Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 1974, 29, 288-298.
- Ekman, P., Friesen, W. V., & Scherer, K. Body movement and voice pitch in deceptive interaction. *Semiotica*, 1976, 16, 23-27.
- Friedman, H. The relative strength of verbal versus nonverbal cues. *Personality and Social Psychology*, 1978, 4, 47-50.
- Mehrabian, A., & Ferris, S. Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, 1967, 31, 248-252.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *Statistical package for the social sciences* (2nd ed.). New York: McGraw-Hill, 1975.
- Scherer, K. R., Koivumaki, J., & Rosenthal, R. Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research*, 1972, 1, 269-285.
- Scherer, K. R., Scherer, U., Hall, J. A., & Rosenthal, R. Differential attribution of personality based on multi-channel presentation of verbal and nonverbal cues. *Psychological Research*, 1977, 39, 221-247.
- Vande Creek, L., & Watkins, J. T. Responses to incongruent verbal and nonverbal emotional cues. *Journal of Communication*, 1972, 22, 311-316.
- Zaidel, S., & Mehrabian, A. The ability to communicate and infer positive and negative attitudes racially and vocally. *Journal of Experimental Research in Personality*, 1969, 3, 233-241.

Received January 5, 1979 ■