Invited article: Face, voice, and body in detecting deceit

Ekman, Paul; O'Sullivan, Maureen; Friesen, Wallace V.; Scherer, Klaus R.

How to cite

EKMAN, Paul et al. Invited article: Face, voice, and body in detecting deceit. In: Journal of Nonverbal Behavior, 1991, vol. 15, n° 2, p. 125–135. doi: 10.1007/BF00998267

This publication URL: https://archive-ouverte.unige.ch/unige:101979
Publication DOI: 10.1007/BF00998267

INVITED ARTICLE: FACE, VOICE, AND BODY IN DETECTING DECEIT

Paul Ekman
Maureen O'Sullivan
Wallace V. Friesen
Klaus R. Scherer

ABSTRACT: Studies based on mean accuracy of a group of subjects suggest that most observers do no better than chance in detecting the lies of others. We argue that a case-by-case methodology, like that used in polygraphy studies may be more useful. Three behavioral measures (two kinds of smiles and pitch) were used to make predictions about the lying or truthfulness of each of 31 subjects. A case-by-case analysis of the hits and misses achieved in this way yielded an over-all accuracy of 86%. The effect on lie detection accuracy of individual differences in the use and control of different behavioral channels is discussed.

Recent reviews of literature on interpersonal deception have attempted to assess which behaviors most powerfully discriminate when someone is lying or truthful (Zuckerman & Driver, 1985; DePaulo, Stone, & Lassiter, 1985). Although these reviews suggested which behaviors yielded consistent results across different studies, as well as statistical estimates of the size of the effects obtained, they did not determine how well any single measure or combination of measures correctly identified when a subject is lying. The statistical approach which has been used—identifying the effects revealed when the entire sample of subjects is considered on a grouped basis rather then determining how many subjects actually showed any difference on each behavioral measure—can not provide such information.

A recent study by Ekman, Friesen, and O'Sullivan (1988) demonstrated the use of a different methodology by determining the relative utility

Paul Ekman's work is supported by a Research Scientist Award from the National Institute of Mental Health (MH 06092).

Correspondence concerning this article should be addressed to Paul Ekman, Human Interaction Laboratory, University of California, 401 Parnassus, San Francisco, CA 94143.

of behavioral discriminators of deceit. In addition to analyzing grouped data (which demonstrated a difference in the occurrence of two kinds of smiling in honest and deceptive interview samples), we also used a data-analysis strategy commonly used in the polygraph lie-detection literature (see reviews by Ekman, 1985; Saxe, Dougherty, & Cross, 1985). We determined how many of the subjects could be correctly classified in terms of when they lied on the basis of each of the two smile measures. Many more subjects could be correctly classified (hits) than not (misses), although about half the subjects did not differ on the smile measures and therefore could not be classified.

We sought, in this further study of these subjects, to determine whether the addition of measures of vocal behavior and body movement would increase the number of subjects who could be correctly classified as lying or truthful. Ekman (1985) proposed that increasing the number of behavioral modalities considered should improve deception detection because of differences among individuals in their ability to monitor and disguise their facial expressions, vocal characteristics or body movements. Based on this reasoning, we hypothesized that there would be more hits (subjects correctly classified as lying or truthful), if measures of either voice or body movement were added to the facial measures.

Method

Deception Scenario

Student nurses were videotaped in a number of standardized interviews. In every interview, the subject watched a short film and answered an interviewer's questions concerning her feelings about it. The interviewer sat with her back to the screen, unable to see what the subject was watching. The subject sat facing the screen and the interviewer. In the first minute of each interview the subject answered questions about her feelings about what she was seeing as she watched the film. Then the film ended, and for the next two to three minutes the interviewer continued to ask questions about the experience. The interviews averaged close to three minutes in duration, with a range from two to almost five minutes. At the end of each interview the subjects rated their emotional reactions to the film and what emotions they thought they had communicated to the interviewer on a nine point scale.

The first two interviews were identified as practice sessions, and the subjects were told they should describe their feelings honestly to the interviewer. The second, supposedly practice, interview was used as the sample of honest behavior. In the third, deceptive interview, subjects saw a

TABLE 1

Means and Standard Deviations of Pleasant and Unpleasant Affect Self-Ratings

	Interview Condition			
	Honest		Deceptive	
	Mean	S.D.	Mean	S.D.
Pleasant Affect Rating				
Felt	7.58	1.45	1.13	.45
Communicated	6.93	1.91	5.27	1.99
Unpleasant Affect Rating				
Felt	1.33	.38	4.44	1.52
Communicated	1.36	.33	2.48	.99

film showing amputations and burns, intended to elicit strong unpleasant emotions. They were instructed to conceal negative feelings and to convince the interviewer they were watching another pleasant film.

The ratings provided by the subjects immediately after the honest and deceptive interviews confirmed that the appropriate emotions were aroused, and that the subjects believed they had communicated what they felt in the honest interview and had *not* communicated what they felt in the deceptive interview. Ratings on happiness and pleasantness were averaged to produce a Pleasant Affect score, and the ratings on anger, fear, disgust, and sadness were averaged to produce an Unpleasant Affect score. A 2 X 2 X 2 ANOVA (Interview Condition X Felt Affect X Communicated Affect) showed that different emotions were felt and communicated in the honest as compared to the deceptive condition, (the three way interaction, $F(2,29) = 172.83$, $p = .001$). Table 1 shows that the subjects reported very pleasant feelings in the honest interview and unpleasant feelings in the deception interview, but the subjects believed they were successful in communicating pleasant feelings in both the honest and deceptive interviews.

This deceptive scenario differs in two ways from the deceptive scenarios used by other investigators. First, the subjects were attempting to conceal very strong emotions felt at the moment of the lie. In most other experiments subjects did not conceal emotions. In the few experiments in

which the lie was about an emotion, it was a mild, not a strong emotion. The second difference is in the motivation of the liars. Our subjects were invited to participate by the dean of the Nursing School and they believed, as we did, that their ability to control their negative emotions while viewing upsetting surgical procedures was important for their later career success. Also, unlike many other experiments on lying, our subjects directly faced the person they were attempting to deceive, and our subjects did not know they were being videotaped until after the experiment was over, decreasing the likelihood that their behavior would be self-conscious.

Our scenario confounds lying and emotion (i.e., being truthful about positive emotion and lying about negative emotion). It provides two samples of behavior which differ in two important ways—the emotion experienced and the veracity of the communication about that emotion. Another limitation in our scenario is that there was no control for the order of the honest and deceptive interviews. The honest interview always came before the deceptive interview, because we found in pilot studies that when the order was reversed the negative impact of the negative films lingered, spilling over into what was intended to be a positive experience in the honest interview. A more detailed description of the scenario was reported by Ekman and Friesen (1974). Ekman, Friesen, O'Sullivan, and Scherer (1980) also provided a more thorough description of the advantages and disadvantages of this paradigm.

Subjects

Fifteen student nurses were recruited via a letter from the Nursing School dean. Two years later another 16 subjects were recruited by the same means, providing a total sample of 31 subjects. The mean age of these women was 20.7, the range was from 19 to 26. After the experiment, all of these subjects reported that it had been helpful in preparing them for their work as nurses, and all of them volunteered when offered an opportunity to go through the experiment a second time.

Behavioral Measures

All of the measurements were made from videotapes or audiotapes. The individuals who scored these behavioral records did not know whether the interviews they scored were honest or deceptive and were unfamiliar with the design or purpose of the experiment.

Face measures. All of the facial behavior shown in the videotaped interviews was measured with the Facial Action Coding System [FACS] (Ekman and Friesen, 1976, 1978). A trained scorer "dissected" each ob-

served expression into the elemental facial muscular actions that produced the facial movement. Further detail about the use of FACS to score these videotapes is provided in Ekman et al. (1988). In that study, Ekman and his colleagues demonstrated that Duchenne or D-smiles occurred more frequently in honest than in deceptive interviews and that masking smiles (Ekman & Friesen, 1982) occurred more frequently in deceptive interviews.

The Duchenne or D-smile occurs when someone is actually experiencing enjoyment. It involves the muscular action of *zygomatic major* and *orbicularis oculi*. The masking smile occurs when traces of the negative emotions a person is trying to conceal are detectable in addition to the smiling action.

Voice measure. The complete utterances of the subjects during the interviews were digitized using a Time/Data 12bit AD converter on a PDP 11/35 minicomputer. The digitized signal was then analyzed with a set of signal-analysis routines developed by Scherer and his associates at the University of Giessen (GISYS, Standke, 1990). Fundamental frequency in Hz was extracted using spectral clipping and zero crossing analysis. This algorithm yields a precision of measurement which compares favorably with other signal-analysis packages (in particular ILS, see Standke, 1990). These values were then averaged over different parts of the interview and the interview as a whole (Guelker, 1977; Stellmann, 1976). An earlier study of 15 of the current 31 subjects found that pitch was higher in the deceptive than honest interviews (Ekman, Friesen & Scherer, 1976).

Body measure. Illustrators is a term Ekman and Friesen (1969) used to distinguish movements which accompany speech and accent, emphasize, trace the rhythm of the speech, or show visually what is being said. Most illustrators are done with the hands, but some can be performed with facial movements. Our illustrator score dealt only with hand movements. Although the type of illustrator shown varies with ethnic background (Efron, 1972), Ekman and Friesen (1972) predicted that illustrators would increase with involvement with the speech process, and decrease when the person is carefully considering each word as it is spoken. In the earlier study of less than half the sample (Ekman et al., 1976), illustrators decreased in the deceptive as compared to the honest interviews.

Results

A repeated-measures MANOVA with condition (honest vs. deception) as the independent factor and the four behaviors as dependent measures indi-

TABLE 2

Univariate Analyses of Honest vs. Deception

Measure	Honest Mean	Honest S.D.	Deception Mean	Deception S.D.	F	p
D-Smiles[a]	2.49	2.03	1.62	1.07	5.90	.022
Masking smiles[a]	.71	1.23	1.14	1.46	6.43	.017
Pitch	220.85	20.58	227.99	23.77	12.76	.001
Illustrators	2.10	2.26	1.46	1.49	2.46	.128

[a]These data are slightly different from those reported in Ekman, et al., 1988, because the sample size is 30, rather than 31, since one subject had unusable pitch data.

cated a significant main effect for condition (F (4,26) = 6.30, $p < .001$). Table 2 gives the means and standard deviations of the four behaviors and the accompanying univariate F-tests. The significant differences between the honest and deceptive conditions for the two kinds of smiles have already been reported and are given here for convenience. The earlier finding with less than half of the current sample, of a significant increase in pitch in deception was replicated with this larger sample ($F(1,29)$ = 12.76, $p < .001$). Although illustrators decreased in deception, as hypothesized, the probability that this difference could have occurred by chance in these data was .128 ($F(1,29) = 2.54$, $p < .128$), which we interpret as a trend.

The subject-by-subject analysis was done by tallying a subject as a hit on a measure if the expected difference (based on earlier findings, Ekman et al., 1976, 1988) between honest and deceptive interviews was found on that measure. Small differences were ignored by requiring that differences in behavior between the honest and deceptive interviews be greater than twice the standard error of measurement. A subject was considered a miss if the difference between the honest and deceptive behaviors was counter to expectation; and unclassified if there was no score for either interview, if the scores were the same, or if the difference was less than twice the standard error of measurement.

Table 3 gives the results of this analysis for the two facial measures, pitch, and illustrators separately, and for some of the combinations of these measures. In the column which reports using both smile measures, a hit was tallied if the difference between the honest and deceptive interviews

TABLE 3

Hits and Misses on the Basis of Facial and Pitch Measures

	D-Smile	Masking Smile	D-Smile + Masking Sm.	Illustrators	Pitch	Smiling + Pitch
Hits	13	11	15	12	18	19
Misses	5	4	5	8	5	3
Unclassified	13	16	11	11	8	9
Hits/ Hits + Misses	72.2**	73.3*	75.0**	60.0	78.3***	86.4***

Binomial Tests: *$p < .059$, **$p < .05$, ***$p < .01$.

was in the predicted direction on both measures, or if it was in the predicted direction on one measure and the other measure had not provided a basis for classification. A miss was tallied if the difference between the two interviews was counter to the predicted direction on both measures, or if it was counter on one measure and the other measure had provided no basis for classification. A subject was tallied as unclassified, if unclassified had been the score for both smile measures, or if one measure yielded a hit and the other a miss. In the column labeled smiles and pitch, the same logic was used except here the two scores used were the classification based on both smiling measures and the single pitch measure.

Table 3 shows that the hit rate with illustrators was low, compared to the face or pitch measures. Because illustrators did not yield significiantly more hits than misses, we did not combine it with either the smile or the pitch measures. Pitch, compared with the smiling measure, provided an additional basis for classifying when subjects were lying or telling the truth. Eight of the subjects who could not be classified on the basis of the two smiling measures were accurately classified by adding the pitch measure. The facial measures, on the other hand, allowed accurate classification of four of the subjects who did not differ in pitch. In addition, there were seven subjects who were correctly classified on both the smiling and the pitch measures, and another seven subjects for whom the smiling measures and pitch measures yielded contradictory classifications: for three subjects smiling yielded the correct classification, and for four subjects pitch yielded the correct classification.

The optimal combination of measures (yielding the highest hit rate) utilized both pitch and the combined facial measures. All three measures—D-smiles, masking smiles, and pitch—gave high hit rates, and no one measure was better than any other (Chi squares comparing each pair of measures were not significant). Table 3 shows that the combination of all three measures yielded the largest number of hits, and the smallest number of misses, although nearly a third of the subjects could not be classified when both the combined smile and pitch measures were used. Combining the smiling and pitch measures yielded results which were significantly different from the results obtained with either of the smile measures considered alone (χ^2 (2, N = 31) = 8.89, $p < .05$ compared with D-smiles; χ^2 (2, N = 31) = 15.35, $p < .001$ compared with masking smiles), but it was not significantly better (although it did discriminate more subjects) than pitch alone or the two smiling measures combined.

Discussion

To our knowledge no prior study of how behavior differs in honest and deceptive interactions has determined the hit rate for each measure and for various combination of measures. Yet this type of subject-by-subject analysis can help to evaluate how well any one behavioral measure or combination of measures can differentiate deceptive from honest behavior as well as in determining whether different behavioral measures are useful in detecting the deception of different people. We hope that other investigators will adopt this type of subject-by-subject analysis, so that it will be possible to compare findings across experiments in regards to which measures allow the identification of lying in the greatest number of subjects.

Duchenne smiles (enjoyment smiles), masking smiles and pitch all yielded high hit rates, although no one measure was better than another. In support of our hypothesis, the hit rate when based on all three measures was very high: 86.4% if the hits are divided by the hits plus misses; 61.2% if the hits are divided by the total sample (misses plus those who could not be classified with these measures). The hit rates found across experimental studies of polygraph lie detection range from 35% to 100%, but the comparison is difficult to make because of major differences in experimental design (see Ekman, 1985, chapter 7, for a review of polygraph studies and a comparison of polygraphic and behavioral clues to deceit). Illustrators did not yield a significant hit rate. We did not utilize any other measures of vocal or verbal behavior in this analysis because measures of hesitations,

speech disruptions, number of words spoken, etc. did not differentiate honest from deceptive behavior in our data (see O'Sullivan, Ekman, Friesen, & Scherer, 1990, for a complete list of the variables measured). Of course there may be other verbal measures, such as statement analysis, which might be more productive.

The finding that about an equal number of subjects could be classified with the face only, the voice only, or both the face and voice, supports Ekman's (1985) suggestion that individuals differ in which aspects of their behavior they may be best able to disguise. Five groups of people can be distinguished: 1) people whose deception was detected with both smiling and pitch; 2) those whose smiles were opaque, but whose pitch increased in deception, as predicted; 3) those whose pitch did not change as predicted but whose deception was revealed by their manner of smiling; 4) people for whom the two measures of smiles and pitch yielded contradictory classifications; and 5) those for whom neither smiling nor pitch provided a basis for classification. Further research is necessary to determine whether such individual differences are stable across different deception scenarios, and if they pertain to the type of behavior (face or voice pitch) which is most emotionally responsive, and therefore most difficult to inhibit, or, instead pertain to the capability for monitoring or controlling one or the other behavior.

Although the hit rate using face and voice measures was quite high, even if these results had been replicated, they should not be applied to any practical attempt to detect lying. Mistakenly judging 16% of the cases would be an unacceptably *high* level if there were any serious consequences of such an identification. Rather than being used as signs of lying, recognition of changes in facial expression and voice might more judiciously be used to alert an interrogator to pursue a line of inquiry.

The hit rate we report is considerably higher than we or others have reported when observers try to judge whether someone is lying without the benefit of behavioral measurement. While most observers do no better than chance, or just slightly better than chance (about 60% correct when chance is 50%), we have found some individual observers whose accuracy reaches 80 to 90% (Ekman & O'Sullivan, in press), and these highly-accurate observers report relying more upon facial and vocal behaviors such as those we measured, than observers who do poorly in detecting deceit.

We believe that a very high level of accuracy in detecting deceit—whether achieved by a gifted observer or through the behavioral measurement we employed in this study—is most likely in deceptions which involve emotion. It is the failure to control completely the involuntary signs of emotion which Ekman and Friesen (1969) postulated generates many of

the behaviors that can betray a lie. Ekman (1985) described two ways in which emotions become involved in deception. 1) Most simply the lie may have as its principal aim the concealment or fabrication of an emotion. 2) Even in lies about non emotional matters, emotions may become involved in the process of lying if the stakes are high (fear of being caught), if the liar and target share values (guilt about lying), or, if misleading the target is considered to be a challenge and an audience is present to appreciate the liar's skill ("duping delight"). Conversely, when the lie is not about an emotion, the stakes are negligible, the target is an unfamiliar, and there is little challenge—which are characteristics found in the lies studied by other investigators—there should be few behavioral clues to deceit, and accuracy should be low. Currently we are experimentally manipulating these characteristics to test this formulation.

Our findings suggest a need to modify Ekman and Friesen's (1969) leakage hierarchy. Their reasoning was that people only control those behaviors which they have learned others pay attention to. Since others rarely comment on the activity of the hands, legs and feet, people will not monitor and disguise these behaviors, but focus more upon facial and verbal behavior. On the basis of the findings from our study and past reviews of the literature, we propose that such a leakage hierarchy may only apply when the motivation to deceive is moderate. When motivation is extremely high, when the stakes for being caught are of major consequence, then most people will monitor and inhibit most of the behavior they are capable of controlling. Verbal content, speech rate and fluency, most body movements, and the large easy-to-see facial expressions, are all more susceptible to deliberate control, we propose, than voice pitch or the type of subtle facial measures which were successful in our study. The merit of this hypothesis awaits testing in studies in which the stakes for success or failure in deception are varied over a wide range.

References

DePaulo, B.M., Stone, J.I., & Lassiter, G.D. (1985). Deceiving and detecting deceit. In B.R. Schlenker (Ed.), *The self and social life* (pp. 323-370). New York: McGraw-Hill.

Efron, D. (1972). *Gesture, race and culture*. The Hague: Mouton. (Originally published in 1941 as *Gesture and environment*)

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, marriage, and politics*. New York: W.W. Norton.

Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49-98.

Ekman, P. & Friesen, W. V. (1972). Hand movements. *Journal of Communication, 22*, 353-374.

Ekman, P., & Friesen, W.V. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology, 29*, 288-298.

Ekman, P., & Friesen, W.V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1*, 56-75.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W.V. (1982). Felt, false and miserable smiles. *Journal of Nonverbal Behavior, 6*, 238-252.

Ekman, P., Friesen, W.V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54*, 414-420.

Ekman, P., Friesen, W.V., O'Sullivan, M., & Scherer, K. (1980).Relative importance of face, body and speech in judgments of personality and affect. *Journal of Personality and Social Psychology, 38*, 270-277.

Ekman, P., Friesen, W. V., & Scherer, K. (1976). Body movement and voice pitch in deceptive interaction. *Semiotica, 16*, 23-27.

Ekman P., & O'Sullivan, M. (In press). Who can catch a liar? *American Psychologist*.

Guelker, E. (1977). *Grundfrequenz und Persoenlichkiet im Interaktionsprozees unter Stress*. Unpublished diploma thesis, University of Giessen, West Germany.

O'Sullivan, M., Ekman, P., Friesen, W.V., & Scherer, K. (1990). *Judging honest and deceptive behavior*. Manuscript submitted for publication.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist, 40*, 355-366.

Standke, R. (1990). *Vocal parameter analysis*. Unpublished manuscript, University of Giessen, West Germany.

Stellmann, H. (1976). Auswirkung von Spannungssituationen auf das Stimmverhalten—Eine Untersuchung dieses Zusammenhangs anhand eines Stressinterviews. Unpublished diploma thesis, University of Giessen, West Germany.

Zuckerman, M., & Driver, R.E. (1985). Telling lies: Verbal and nonverbal correlates of deception. In A. W. Siegman & S. Feldstein (Eds.), *Multichannel integration of nonverbal behavior* (pp. 129-147). Hillsdale, NJ: Erlbaum.