Frank Horvath,[1] *Ph.D.; Brian Jayne,*[2] *M.S.D.D.; and Joseph Buckley,*[2] *M.S.D.D.*

Differentiation of Truthful and Deceptive Criminal Suspects in Behavior Analysis Interviews

REFERENCE: Horvath, F., Jayne, B., and Buckley, J., **"Differentiation of Truthful and Deceptive Criminal Suspects in Behavior Analysis Interviews,"** *Journal of Forensic Sciences*, JFSCA, Vol. 39, No. 3, May 1994, pp. 793–807

ABSTRACT: The Behavior Analysis Interview© (BAI) is a commonly used procedure designed to assist investigators in distinguishing between suspects who are concealing their involvement in a criminal event (deceptive) from those who are not (truthful). During a BAI a protocol of questions is asked and suspects' verbal responses and accompanying nonverbal behaviors and attitudinal characteristics are assessed. Based on this assessment the likelihood of involvement in the criminal event is determined.

The purpose of this study was to determine the effectiveness with which trained evaluators were able to distinguish between truthful and deceptive suspects undergoing BAIs. Sixty videotaped interviews, 30 of truthful and 30 of deceptive suspects, were observed by four evaluators, each of whom independently scored suspect's behaviors and attitudes and judged the suspect's truthfulness. The results showed that, excluding inconclusive decisions, evaluators' average accuracy on truthful suspects was 91% and on deceptive suspects, 80%. Suspects' status did not affect confidence of evaluators' decisions but confidence was greater when correct as opposed to incorrect calls were made. Deceptive suspects manifested "theoretically" predicted behaviors and attitudes of "deceptiveness" to a significantly greater degree than did truthful suspects. The BAI appears to be useful for investigative purposes in order to differentiate between suspects who are concealing involvement in a criminal offense from those who are not.

KEYWORDS: criminalistics, behavior, suspects, interviews

In spite of the popular, and sometimes even professional perception to the contrary, it is relatively uncommon to find that criminal investigations are resolved solely because of systematic sleuthing and scientific successes [*1–5*]. In fact, physical evidence and scientific analyses become most useful once a suspect is identified, although they can, of course, be helpful in circumscribing investigative efforts. Nevertheless, it is interpersonal communication that most typically leads to the identification of an offender. Simply put, detectives "solve" cases by talking with victims, witnesses and suspects and, in many instances, by interviewing and interrogating criminal suspects. In performing these

Received for publication 19 July 1993; revised manuscript received 22 Oct. 1993; accepted for publication 23 Oct. 1993.

[1]Professor, School of Criminal Justice, Michigan State University, East Lansing, MI.

[2]Director of Research and Development and President, respectively, J. E. Reid and Associates, Chicago, IL.

Presented at American Academy of Forensic Sciences, February 1993, Boston, MA.

communication tasks detectives are almost always confronted with an age old dilemma: how to distinguish between those who are telling the truth and those who are not [*6,7*]. How they make such distinctions and the degree to which they are successful at it is likely to depend on a number of personal and professional qualifications and the extent to which there are observable differences between those who tell the truth and those who lie.

Although interrogation and interviewing serve different goals and are related but separate procedures, the questioning of criminal suspects, whether to gain information (interview) or to elicit an admission against interest (interrogation) are essential practices in which successful detectives spend a large portion of their time [*2,3,5*]. Perhaps, in recognition of this, field practitioners in recent years have developed methodologies of questioning suspects that are said to improve the ability to distinguish between those who are concealing involvement in a criminal offense ("deceptive") from those who are not ("truthful"). These methods rest on the assumption that there are observable behavioral differences between deceptive and truthful persons. These differences may lie in verbal, nonverbal, paralinguistic and attitudinal dimensions, many of which are widely discussed in the popular literature [*8,9*] but not strongly substantiated by rather extensive scientific investigation [*10–19*].

There is a great deal of disagreement in the scientific literature about the usefulness of specific behaviors as indicators of deception, although there is agreement that at least for some measures, observers can detect lies (or liars), or at least certain statements of "liars," at rates slightly exceeding chance levels [*12,17,20*]. It is frequently mentioned, however, that one of the serious shortcomings in the available research is that almost all of it has been carried out in laboratory environments [*21*]. In such circumstances it is difficult, if not impossible, to replicate the emotionality, the motivation and the psychological orientation that would be expected in real-life circumstances. When one's reputation, employment, or the possibility of arrest and prosecution are at stake, behavioral indicators of deception may differ in degree, if not in kind, from those that are observed in artificially constructed laboratory settings.

Although the research that is not laboratory based is sparse there are four reports that shed light on behavioral differences between truthful and deceptive persons in real-life settings. In the first of these Reid and Arther [*22*] showed that when undergoing polygraph testing, liars tended to exhibit certain mannerisms such as, poor eye contact, nervousness, etc., not generally displayed by truth-tellers. In addition, liars more often expressed reservations about the situational context and were more likely to complain about the procedure than were truth-tellers. These observations were supported by Horvath [*6*] who tabulated verbal and nonverbal behaviors of a sample of 100 suspects who had undergone polygraph examinations; the behaviors of the deceptive suspects were generally significantly different from those of truthful suspects. Unfortunately, in both of these studies there was a reliance on information collected impressionistically and, while these data are certainly suggestive of behavioral differences between truthful and lying suspects, there is reason to be cautious in generalizing from them.

A third field study was reported by Barland [*23*] as part of a larger project to determine the validity of polygraph examinations. Barland scored a small sampling of suspects' behavior and found that 87% of his behavior-based assessments correctly predicted test outcomes on guilty (deceptive) suspects whereas only 50% of the assessments on innocent (truthful) suspects were predictive of the polygraph outcomes. Barland also reported, however, as some field polygraph examiners contend, that his assessment of suspects' behavior provided an important reality check on the polygraphic data.

The fourth and most recent field study by Horvath and Jayne [*24*] investigated the relative contribution of different sources of behavioral data derived from structured interviews of persons suspected of theft. Four trained evaluators made judgments of 14 confession-verified deceptive and six confession-verified truthful suspects' behavioral

responses to a set of standard interview questions. These judgments were made under four different conditions. In three of these conditions the evaluators' judgments were based on an assessment of the interviewer's questions, and the suspects' responses to them, when each of the question/response segments was considered independent of the others and out of the context in which they originally occurred. In one of these conditions, judgments were made solely by review of a written transcript of the question/response segments; in another an audio recording was reviewed; and, finally, in another condition an audiovisual recording of the question/response segment was evaluated. In most of these evaluations the accuracy of the decisions made by the evaluators exceeded chance levels but was not exceptionally high. However, when evaluators made judgments in a fourth condition, by reviewing question/response segments in the context in which they actually occurred, the accuracy of the decisions was in excess of 90% on both truthful and deceptive suspects.

These results suggest that structured interviews may produce findings quite different from those typically reported in the laboratory [*21*]. However, the Horvath and Jayne [*24*] study used only a small sample and all of the interviews that were assessed had been verified by a confession of the guilty person in each investigation. Consequently, the suspects may not have been a representative group. That is, as is true with other field research involving detection of deception, suspects in confession-verified cases may differ from the population encountered in most real-life circumstances [*25*]. In addition, this study, as well as all of the other field research, relied on behavioral assessments of suspects involved in polygraph-testing circumstances. Since most police questioning of suspects is not carried out in that context, the research may not generalize to nonpolygraph settings.

From this brief review, it can be seen that while the available literature provides only sketchy support for a correlation between behavioral cues and deception, the direction and nature of the information suggests that in real-life circumstances the value of behavioral assessments may differ from that commonly reported in laboratory-based studies. This conclusion is reinforced by the development and widespread application in recent years of "behavior analysis interviewing," a method of discriminating between "truthful" and "deceptive" suspects based on observations of their behavior while undergoing a structured interview. This technique, the "Behavior Analysis Interview" (BAI) was originated by J. E. Reid and Associates [*26*]. The BAI was empirically developed from experiences in interviewing and interrogation and although it was initially carried out in the context of polygraph testing, it is now performed independent of that context and does not involve instrumental (that is, physiological) measurements.

Generally, investigators who have been trained in the BAI technique have reported quite favorable results. Moreover, there is an increasing number of field personnel who seek out such training. These observations indicate that sound research on the effectiveness of the BAI would be of practical as well as scientific value. This paper reports the results of a project designed to investigate—as a part of a larger study—the accuracy of classifications made on the basis of the BAI in real-life circumstances.

The Behavior Analysis Interview

In this study, all suspects whose behaviors were assessed freely chose to practice deception or to tell the truth, knowing that there were likely serious consequences for failing to avoid detection or for giving false indicators of deception. Because these suspects were questioned during BAIs is necessary that the general nature and format of these interviews is understood. Therefore, a brief description of the BAI follows.

The BAI consists of a 30 to 45 minute non-accusatory, structured forensic interview designed to elicit verbal and nonverbal behaviors and attitudinal characteristics of the

suspect being questioned. During the initial period of the BAI the interviewer (investigator) seeks background information from the suspect and attempts to establish "normative" behavioral patterns such as "eye contact," response latency and nervousness. Questions asked after behavioral norms are established are either those that attempt to assess the suspect's opportunity, motivation and propensity for involvement in the issue at hand ("investigative" questions) or, are those used to elicit differential verbal and nonverbal behaviors and attitudinal characteristics from truthful and deceptive persons. Questions used for the latter purpose are known as behavior-provoking questions.

These questions have been developed empirically and models useful for interpreting the content and the accompanying verbal and nonverbal behaviors of suspects' responses to them can be found in the literature [*6,26,27*]. However, since there has been little information reported about truthful and deceptive suspects' attitudinal and related differences during the asking of such questions, a brief overview is given here.

During the asking of behavior-provoking questions, deceptive suspects reveal an attitude toward the investigation in which they are involved, and the interview about it, that differs from that manifested by truthful suspects. Deceptive suspects, for example, are said to be less likely to offer helpful investigative information to an interviewer. They do not exhibit an appropriate level of concern about being a "suspect" and often lack spontaneity and sincerity in their responses. They speak in a guarded way and appear to edit their verbal responses. On the other hand, truthful suspects are helpful to the interviewer and show an expectancy to be exonerated. They often exhibit resentment toward the "guilty" person and their responses to the interviewer's questions are spontaneous and sincere.

The evaluation of a suspect's attitude may involve consideration of a number of behavioral cues, including postural changes. A more open, forward-leaning and comfortable posture, for example, is indicative of "truthfulness" and a positive attitude, whereas a rigid, frozen and defensive posture is commonly associated with "deception."

Aside from the report of Horvath and Jayne [*24*] there has not been any systematic research on the effectiveness of the BAI in discriminating between truthful (innocent) and deceptive (guilty) persons. Since, the BAI is now rather widely used in the investigation of criminal offenses, such an assessment is reported here.

Method

Data Collection

This study required the analysis of behaviors drawn from real-life settings but, because it was not possible to do such analysis contemporaneous with the occurrence of the behaviors, it was necessary to collect audiovisual (AV) documents from interviews of persons being questioned about involvement in criminal activity. This was done by videotape recording BAIs conducted by trained and experienced staff members of John E. Reid and Associates. The selection of the sample of these AV documents and the procedures used in subjecting them to analysis are described in this section.

Sample Selection Criteria

In order to select a sample of AV documents as free from selection bias as possible, a procedure was established to obtain a sample of AV documents prospectively. It was desired, however, that the sample meet certain criteria for selection. These were: (1) that each BAI involve an investigation of a loss or suspected theft of a specific amount of money or property. BAIs that investigated an on-going loss or general inventory shortage or an offense not involving theft were excluded from consideration; (2) that the gender,

race, type of verification (that is, either confession or other corroboration of ground truth) and verification status (that is, truthfulness or deception) were as evenly matched as possible in the final sample; (3) that BAIs conducted by those who were assigned to serve as evaluators in the study were excluded from consideration; and, (4) that "ground truth," that is, the actual truthfulness or deception of the suspect, could reasonably be established.

Between November 1, 1989 and November 15, 1991 110 BAIs carried out on the premises of J. Reid and Associates were made available for AV-taping. This group included all BAIs carried out during that period except for those that had to be excluded because a suspect declined to be video-taped, equipment was not available, or the issue did not involve theft. Of these 110 BAIs, 23 (21%) had to be excluded either because the interviewer was scheduled to participate in this study or the interview itself concerned multiple thefts that may have been unrelated.

Thus, 87 videotape documents, each an AV recording of one person undergoing a BAI, were ultimately available for use in the study. These 87 documents represented BAIs carried out in 56 independent theft investigations which interviewers employed by J. Reid and Associates were hired to investigate. Five different interviewers administered these BAIs, each specifically trained in behavior analysis techniques. The experience of these interviewers in administering BAIs averaged ten years with a range of from 8 to 14 years.

Establishment of Ground Truth

Of the 87 suspects' whose interviews were video taped the "guilt" (deception) or "innocence" (truthfulness) of 34 of them was established by a corroborated confession of the guilty suspect in each of the investigations in which these persons had been involved. In two other instances, the suspects' innocence was established by solid information that showed that the theft under investigation did not actually occur. Hence, the ground truth status of 36 suspects was confirmed by either a confession or, in two instances, by other documented circumstances that verified truthfulness. For simplicity, unless otherwise specified, all of these are referred to as "confession-verified" cases in the remainder of this paper. The remaining 51 suspects were not involved in either confession-verified cases or in cases in which there was other reasonably certain independent verification. For that reason, "ground truth" was established in these cases by means of systematic factual analysis as described by Jayne [*28*].

To carry out factual analysis it was necessary to process the information pertaining to each of the different investigations as the audiovisual tapes in each were collected. To do this case facts were extracted from each case file as well as from information provided by the suspect(s) during the interviews. These data were summarized in a narrative statement that was then given to two evaluators who independently completed a "factual analysis data sheet" on each suspect.

Factual analysis required the two evaluators, neither of whom was otherwise involved in the study, to assign individual probabilities of "guilt" or "innocence" in each of five separate areas for each suspect. These were: biographical; opportunity/access; personal activities; motivation/propensity; and, evidence.

From the two evaluators' results in each of the areas of evaluation an overall probability of guilt or innocence for each suspect was calculated by one of the researchers (BJ). The evaluators who completed factual analysis did not know whether their assessments, individually or collectively, would include or exclude a suspect from the study group.

All 87 suspects were subjected to factual analysis. Of the 36 suspects whose ground truth was established by confession or other evidence, only one produced a final score

from both evaluators which was greater than 90% and that was also inconsistent with ground truth. That is, when there was agreement by both evaluators at levels of 90% or higher all but one of the confession-verified suspects were correctly classified. Therefore, requiring at least a 90% confidence level from both evaluators for the inclusion of non-confession-verified suspects appeared to provide a satisfactory criterion when ground truth could not otherwise be established.

From the group of 87 video taped BAIs, 60 were included in the sample. These included all 34 suspects (13 truthful and 21 deceptive) whose interview results had been verified by a confession. Two other truthful suspects were included because information was developed that independently established their truthfulness. The remaining 24 suspects in the sample (15 truthful, 9 deceptive) were included because: (1) factual analysis by two independent evaluators indicated at least a 90% probability of truthfulness or deception; and, (2) they most closely balanced the sample in terms of "truthfulness" (status), race and gender. In this way, 60 video tapes of field BAIs were assembled for use in the study.

Table 1 shows the number of suspects in the sample categorized by gender, race, and type of verification. It can be determined from those data that 57% of the suspects were female; 43% were male. Fifty-five percent of the suspects were white and 45% were non-white. Sixty percent of the sample had been verified by confession, 40% by factual analysis.

Differences Between Sample Subjects

The mean age of all suspects in the sample was 28 years (s = 8.22). White suspects had a mean age of 27.1 years (s = 8.5) while the mean age of non-white suspects was 28.0 years (s = 7.9). Females had a mean age of 29.7 years (s = 8.8); the mean age of males was 24.7 years (s = 6.3). An analysis of variance (Anova) in which there were three factors (Status—truthful/deceptive; Race—white/non-white; Gender—male/female) with age as a dependent variable indicated that there were no statistically significant differences in the mean age of the sample for Race or Gender. However, the mean age of those who were deceptive (M = 24.4) was significantly [$F(1,52) = 5.8$, $P < .01$] lower than that of suspects who were truthful (M = 30.6). The mean age of suspects who were confession-verified was 26.4 years (s = 6.8); for suspects whose status was established through factual analysis the mean age was 29.2 years (s = 9.8). A t-test showed that this difference was not statistically significant [$t(59) = -1.3$, $P = .21$]. Chi-square tests revealed no significant relationship between the method of establishing ground truth (con-

TABLE 1—*Distribution of sample suspects by method of verification, race and gender.*

	Verification Type			
	Confession		Fact Analysis	
	(n = 36)		(n = 24)	
Race/Gender	Truthful	Deceptive	Truthful	Deceptive
Non-White/				
Female	4	5	5	3
Male	2	4	1	3
White/				
Female	4	6	7	0
Male	5	6	2	3
Totals	15	21	15	9

fession/factual analysis) and either Race [$\chi^2(1) = .40$, $P > .05$] or Gender [$\chi^2(1) = .55$, $P > .05$].

As shown in Table 1, the sample included males and females and white and non-white suspects. Because it was not possible to balance the sample such that these characteristics were equally distributed and, since the interest here focused on differences between truthful and deceptive suspects, these variables were not considered in statistical analysis.

Preparation of Audio-Visual Documents

Each of the 60 AV tapes contained the complete BAI (30 to 45 minutes) of one of the 60 suspects included in the sample. Each of these BAIs originally included three types of questions asked by the interviewer: background ("normative") questions, which established the suspect's age, marital status, employment position, salary and related items; "investigative" questions that gave the suspect an opportunity to provide information about the investigation at hand; and finally, "behavior-provoking" questions, asked to elicit behavior to differentiate between truthful and deceptive suspects [*6,26*].

From each of the 60 video-taped interviews, the 15 behavior-provoking questions that are the most commonly asked during theft investigations were extracted and dubbed onto a single tape for each suspect. The interviewer's asking of each question as well as the suspect's response (and accompanying behavior) were included in each dubbed segment. However, even though it is customary in real-life BAIs for interviewers to ask "follow-up" questions when a suspect's initial response to a behavior provoking question is ambiguous, such follow-up questions were excluded in order to maintain greater consistency between each of the 60 video documents. Moreover, in each of the dubbed video documents that were prepared, the order of the behavior-provoking questions was identical for all suspects even though in actual field conditions the order may vary somewhat. Finally, it is important to note that in real-life circumstances, an interviewer may consider it inappropriate to ask certain questions because of the investigation at hand; for that reason, the number of behavior-provoking questions was not identical for each suspect.

In addition to the 15 behavior-provoking questions, a 90-second block of each suspect's behavior during the asking of "normative" questions was also extracted. The normative behavior segment for each suspect was dubbed onto a video tape for each of the 60 suspects. In this way a single videotape document was prepared for each suspect; the normative behavior was presented first in sequence on this tape; that segment was followed by the behavior-provoking questions that were available for each suspect presented, as specified previously, in the same order on each tape.

The behavior-provoking questions extracted for use in this study in the order in which they were presented on each of the 60 video tapes were as follows [see: *6,26,27*]:

1) *Purpose*:
What is your understanding for the purpose of this interview today?
2) *You*:
(Name) If you stole (this money) you should tell me that now. Did you steal that money?
3) *Knowledge*
Do you know who stole (this money)?
4) *Suspicion*
Who do you suspect may have stolen (this money)?
5) *Vouch*
Is there anyone you can vouch for, who you do not think was involved in (this theft of money)?

6) *Opportunity*
Who would have had the best opportunity to (steal this money) if they wanted to?
7) *Think Stolen*
Do you think this (money) was actually stolen?
8) *Feel*
How do you feel about being interviewed regarding this (theft)?
9) *Results*
How do you think the investigation will come out on you?
10) *Think*
Have you ever thought about (stealing money)?
11) *Punishment*
What do you think should happen to the person who stole (this money)?
12) *Second Chance*
Do you think the person who (stole this money) should be given a second chance?
13) *Why Not*
Tell me why you wouldn't (steal this money)?
14) *Motive*
Why do you think someone did (steal this money)?
15) *Tell Loved One*
Have you told your (mother/spouse/family) about coming in for the interview today?

Evaluation of Videotape Documents

Four persons, trained and experienced in "Behavior Analysis Interviewing," viewed each of the 60 videotape documents over the course of a number of monitored reviewing sessions. In these sessions the evaluators independently blind-scored several dimensions of suspects' behavior and made judgments of their status. These included the following:

(1) What was the evaluator's opinion of the suspect's attitude on the following dimensions: "Insincere," "Unconcerned," "Unhelpful," and "Guarded." Judgments of each of these were scored on a five-point scale, ranging from "1" to "5." A value of (1) on this scale indicated "Definitely not" and a (5) indicated "Definitely yes." Hence, lower values were associated with a lesser degree and higher values a greater degree of the "attitude," with higher scores indicating a better fit with theoretical "models" of deceptive persons' attitudes. In this scoring, evaluators were permitted to score each attitude with a zero (0) to indicate that they were unable to make a judgment. This was provided because it was recognized that some of the behaviors may not have been sufficiently distinct to permit a definitive judgment.

(2) What was the evaluator's assessment of the suspect's posture in each of the following areas: "Closed," "Uncomfortable," and "Rigid/Frozen." Judgments of each of these were scored on a five-point scale, ranging from "1" to "5." A value of (1) on this scale indicated "Definitely not" and a (5) indicated "Definitely yes." Again, lower values were associated with a lesser degree and higher values a greater degree of the postural item; higher scores indicated a better fit with "deceptiveness." As in other similar evaluations, evaluators were permitted to score each posture with a zero to indicate that they were unable to make a judgment.

(3) What was the evaluator's overall opinion as to whether or not the suspect's behavior indicated that the suspect was a "truthful" person, a "deceptive" person or that it was not possible to determine the status, that is, was "inconclusive."

(4) What was the evaluator's degree of "confidence" in his decision. This was indicated on a six-point scale, anchored at the low end with "none" (0) and "Very high" (5) at the opposite end.

Results

Unless otherwise noted, all statistical testing used the .05 level as the criterion for statistical significance. Also, the probability levels reported are for one tailed tests unless otherwise noted.

It will be recalled that there were two different methods of establishing ground truth for the audiovisual documents included in the study, by confession or other solid evidence and factual analysis. Because these two methods differ considerably, it was important to determine whether or not this difference would affect subsequent analyses. For that reason, evaluators' scores on both the confession and the fact-analysis verified suspects were compared. This was done by separately subjecting each evaluator's scores for the various assessments of suspects' behavior to a Two-way Anova in which Status (truthful/deceptive) and Verification (confession/facts) were the two factors. These analyses did not reveal any significant effects for the Verification factor. Moreover, Chi-square tests showed no relationship between Verification status and the frequency of correct, wrong and inconclusive judgments. Therefore, the type of verification was not included as a variable in further statistical calculations.

After viewing each of the 60 BAIs, each evaluator indicated his decision of truthfulness and deception, his confidence in each decision and his rating of the suspects' attitudes and posture. Attitudinal items that were scored were: "sincerity," "concern," "helpfulness," and "guarded"; postural items included "closed posture," "comfortableness," and "rigid."

The distribution of evaluators' decisions by suspects' status is shown in Table 2. It can be seen in that table that when averaged across the four evaluators, 78% of the judgments on actually truthful suspects were "truthful" decisions, 8% were "deceptive" and 14% were "inconclusive." On deceptive suspects, the evaluators' averaged 66% "deceptive" decisions, 17% "truthful" and 17% "inconclusive." Thus, there were generally more errors made on deceptive suspects, than on truthful suspects; that is, there were more false negatives than false positives.

In actual investigations "inconclusive" decisions usually result in additional investigation of a suspect. For this reason, it is inappropriate to view these decisions as errors. Consequently, the accuracy of the evaluators' decision was calculated by excluding inconclusive outcomes. When this was done, the evaluators' accuracy on truthful suspects

TABLE 2—*Percent and number of truthful deceptive and inconclusive decisions for each evaluator by suspects' status.*

Suspect Status/ Decision	Evaluator A % (*n*)	B % (*n*)	C % (*n*)	D[a] % (*n*)	Mean %
Truthful/					
Truthful	83 (25)	80 (24)	77 (23)	73 (22)	78
Inconclusive	7 (2)	7 (2)	23 (7)	20 (6)	14
Deceptive	10 (3)	13 (4)	0 (0)	7 (2)	8
Deceptive/					
Truthful	20 (6)	20 (6)	10 (3)	17 (5)	17
Inconclusive	23 (7)	3 (1)	23 (7)	20 (6)	17
Deceptive	57 (17)	77 (23)	67 (20)	63 (19)	66

[a]Percentages shown in each cell were calculated separately for decisions on Truthful and Deceptive suspects for each evaluator. Chi-square tests, corrected for continuity, calculated on each evaluators' decisions, excluding inconclusives, and suspects' actual status were, for evaluators A through D, respectively: $\chi^2(1) = 18.6$, Phi = .64; 21.6, Phi = .65; 31.9, Phi = .88; 21.7, Phi = .71.

TABLE 3—*Correlations between pairings of evaluators on decisions of truthfulness and deception.*

	Evaluator			
Evaluator	A	B	C	D
A	...	.71	.87	.58
B	...	...	.60	.63
C	...	...	...	.72

ranged between 89% and 100%, with a mean accuracy for the four evaluators of 91%. On deceptive suspects the evaluators' accuracy ranged between 74% and 87% with a mean of 80%. The mean accuracy across all evaluators and all suspects was 86%.

Further statistical analysis was performed to examine the relationship between evaluators' decisions and suspects' actual status (when inconclusive judgments were excluded). These analyses showed statistically significant relationships for all four evaluators. The strength of these relationships, determined with the Phi coefficient, was calculated for each evaluator; as shown in Table 2, Phi was .64, .65, .88 and .71, for evaluators A, B, C, and D, in order.

To determine evaluators' agreement in judgments of truthfulness and deception, each of the six possible pairs of evaluators' decisions were compared. The percentage of instances in each pair when the two evaluators reached a common decision (not counting inconclusive judgments) ranged between 86% and 98% with a mean of 89%. In addition, each evaluator's decisions across all 60 subjects were correlated with those made by each of the other evaluators. These correlations, calculated, of course, with inconclusive decisions included, are shown in Table 3. As indicated, they ranged between .58 and .87 for the six pairings of evaluators.

It will be recalled that evaluators indicated their confidence in their decisions and also scored each suspect with respect to certain attitudinal and postural characteristics. The mean scores calculated across all four evaluators for each of these characteristics, separately indicated for truthful and deceptive suspects, are shown in Table 4. (Scores of "0" on the behavioral measures, indicating an inability to make a judgment, were excluded from the calculations.) Here it can be seen that evaluators' confidence was not

TABLE 4—*Means and standard deviations for evaluators' scorings of confidence and of suspects' behavior and attitudes.*

	Truthful (n = 30)		Deceptive (n = 30)		
Measurement	Mean	SD	Mean	SD	t(58)=
Evaluator Confidence	3.28	.70	3.35	.85	−0.3 n.s.
Suspects' Attitudes:					
Sincerity	1.95	.44	3.17	.81	−7.3[a]
Concern	1.90	.46	2.92	.83	−5.9[a]
Helpfulness	1.94	.57	3.32	.80	−7.7[a]
Guarded	2.28	.68	3.63	.78	−7.2[a]
Suspects' Posture:					
Closed	2.73	.94	3.48	.97	−3.0[a]
Comfortableness	2.43	.53	3.38	.73	−5.6[a]
Rigid	2.69	.58	3.33	.72	−3.8[a]

[a]Significant difference between truthful and deceptive suspects, 1 tailed tests, $P < .001$.

significantly different for decisions made on truthful (M = 3.3) and deceptive suspects (M = 3.3) [t(58) = −.33, $P > .10$]. That is, they were equally confident in their decisions regardless of suspects' actual status. Additional analysis of confidence scores showed that the mean confidence score on correct decisions was significantly higher than on incorrect decisions, excluding inconclusives, for three of the four evaluators. For evaluators A through D, in order, the mean scores on correct and incorrect decisions were: 3.6 and 2.8 [$t(49) = 2.0$, $P < .01$]; 3.9 and 2.5[$t(55) = 3.6$, $P < .01$]; 3.3 and 2.3[$t(44) = 1.7$, n.s.]; 3.3 and 2.0 [$t(46) = 2.7$, $P < .01$]. Hence, evaluators' were generally more confident when they judged suspects' status correctly than when they were wrong.

Also shown in Table 4 are the mean scores for the attitudinal and postural characteristics on both groups of suspects. Here, it can be seen that deceptive suspects' mean score on each of the items was significantly greater than that for truthful suspects. For example, deceptive suspects' average score on "sincerity" was 3.17 (S = .81) whereas truthful suspects averaged 1.95 (S = .44); this difference was statistically significant and indicated that deceptive suspects were seen to display the (lack of) "sincerity" believed to be characteristic of actually deceptive persons to a greater degree than did the truthful suspects. Similarly, deceptive suspects were found to be significantly more apt to show a "closed posture," more characteristic of those who are concealing the "truth" than those who are not; the mean scores here were 3.48 (S = .97) and 2.73 (S = .94) for deceptive and truthful suspects, respectively. In other words, for all of the measures of suspects' attitudes and posture deceptive suspects were seen to show more characteristic "deceptive" behaviors than truthful suspects.

Because the behavioral items produced similar effects, inter-item correlation coefficients were calculated on the (four evaluators' average) scores for each item across all 60 suspects; these are shown in Table 5. In that table the items have been grouped according to whether they reflected an "attitudinal" behavior or one that showed a dimension of the suspects' posture. The former category included concern, guarded, helpful and sincerity; the latter, closed, comfort and rigid. It can be seen in Table 5 that the attitude items were strongly inter-correlated, with coefficients ranging between .85 and .90. The items in the posture category were less highly correlated with each other, with r ranging between .37 and .67. The r values between the posture items and those in the attitude category ranged between .39 and .81.

Given the strong inter-correlations of the behavioral items, it was decided to merge all "attitudinal" items into one measure; this was done by calculating the mean value for all evaluators across all four of the attitude items to produce a single "attitude" score. A similar measure, a "posture" score was calculated on the three "posture" items. These attitude and posture scores were then separately subjected to statistical analysis.

The result of this analysis showed that the mean attitude score on truthful suspects, 2.0 (s = .50), was significantly smaller than that on deceptive suspects, 3.3 (s = .73),

TABLE 5—*Inter-item correlations for seven behaviors of suspects rated by evaluators.*

Item	Attitudes				Posture		
Attitudes	C	G	H	S	Cd	Ct	R
Concerned (C)	...	.85	.87	.88	.57	.70	.45
Guarded (G)		...	.90	.90	.67	.81	.54
Helpful (H)			...	.89	.54	.69	.49
Sincerity (S)				...	.53	.73	.39
Posture							
Closed (Cd)					...	.67	.37
Comfort (Ct)						...	.39
Rigid (R)							...

[$t(58) = 7.6$, $P < .001$]. Analysis of the posture score also revealed a statistically significant effect, with a mean score of 2.6 ($s = .52$) on truthful suspects and 3.4 ($s = .62$) on deceptives [$t(58) = 5.3$, $P < .001$]. Thus, the attitudes and postures of the suspects were consistently in the direction predicted by the empirically developed model on which the BAI is based.

Discussion

These findings suggest that when questioned in a structured Behavior Analysis Interview the behaviors of suspects who attempt to conceal their involvement in serious acts of wrongdoing vary in many respects from those who do not. The behaviors of deceptive suspects, in other words, differ from those of truthful suspects. These behavioral differences are detectable in the "attitudes" displayed toward the issue under investigation and the interview process, and in postural as well as possibly other behaviors. However, it is important that we point out here that to be meaningfully evaluated, behaviors such as those considered here must be evaluated in the context of the setting in which they occur. This implies, of course, that behavior analysis as it is practiced in the BAI, involves the evaluation of substantially more behavioral information, and more types of information, than those which are commonly the subject of research on the relationship between deception and behavior [*10,12*].

It is also important to point out that in actual Behavior Analysis Interviews the interviewer has access to not only all of the behavioral data considered in this study, but other information as well. That is, in an actual interview, the suspect, in addition to being asked behavior-provoking questions, would also be asked investigative questions to ascertain motivation, propensity, and opportunity to commit the offense. During the BAI, the interviewer also has control over the selection, sequence, and timing of questions, as well as the ability to ask follow-up questions to clarify ambiguous responses. Moreover, the interviewer often has access to information about the suspect's background, as well as the suspect's role in the investigation. This information, of course, may assist in evaluation of behavior and allow the interviewer to estimate independently the probability of the suspects' involvement in the crime. While the contribution of these additional sources of information to an interviewer's decision-making is unknown, it can be seen that the range of data, behavioral and otherwise, that is available in an actual BAI may contribute differently in real-life from what has been shown here.

There are three other important points to be kept in mind when considering these findings. First, all of the expert evaluators in this study were well trained and highly experienced in interpreting the kind of behavioral information submitted to them. Whether or not others with lesser or different qualifications would do as well is simply not known. In fact, recent research of Ekman and O'Sullivan [*21*] suggests that even among those with a professional interest in the detection of liars from behavioral differences, such as police detectives, there may be considerable variation.

A second point is that the clarity of behavioral responses exhibited by a suspect during an interview may be influenced by the interviewer's ability to evoke such behaviors. That is, it is possible that important features in the interaction between an interviewer and the interviewee are determinative of at least some of the behavioral data studied here. Other interviewers who are less able to elicit meaningful behavior from suspects might not produce similar kinds or amounts of behavioral information.

Finally, although these results show that there are observable behavioral differences between those who tell the truth and those who do not the specific behavioral features that contribute to those differences are not certain. Further research is necessary to reveal the nature and source of these cues.

In spite of these limitations, it is true, nevertheless, that the behavioral data evaluated

here were derived from actual suspects involved in investigations of serious wrongdoing. This feature alone, as indicated at the outset of this report, distinguishes these findings from most other research. When considered in that light, the Behavior Analysis Interview appears to be empirically well grounded. Although other research has reported only a moderate discrimination between truthful and deceptive persons based on behavioral observations [*10,12,29*], almost all of this research has been carried out in a laboratory environment, usually in conditions in which the subjects had little to gain by telling the truth and little to lose by lying. In contrast, in this study the subjects were involved in real-life investigations; they knew that there were serious consequences for failing to be identified as truthful. Moreover, here the behaviors that were assessed were those occurring in interviews specifically designed to elicit behaviors and attitudes believed to discriminate between truthful and lying suspects. Judging from these results such interviews are quite effective and yield a relatively high overall accuracy of about 85%. Because these data were real-life based, however, there is an important qualification to be made.

Unlike laboratory studies, in field settings it is extremely difficult to develop an adequate measure of "ground truth," that is, a criterion that establishes with certainty the "innocence" or "guilt" of a suspect that is truly independent of the process being evaluated. This problem complicates the interpretation of field, or real-life, based research such as the present study. In this study, for example, two primary measures of ground truth were used: confessions, which implicated a "guilty" suspect and at the same time exonerated an "innocent" suspect in the same case, and factual analysis, a method to determine the probability of involvement (or the lack of it) in an offense based on information compiled about a suspect's activities related to the offense but without any direct observation of the suspect or his/her behavior mannerisms.

In this study, the interviewers, who administered the actual BAIs in the confession-verified cases, no doubt considered information that was not available to the evaluators, such as investigative data, when rendering their field opinions. The evaluators, however, did not have access to such data and were forced to rely only on their assessment of responses to the behavior-provoking questions asked during a BAI. Because both the interviewers and the evaluators had access to the same behavioral information, however, it could be suggested that the congruence between the outcome of the original BAI and evaluators' assessment of the suspects' truthfulness, reveals only a measure of the consistency between the original and subsequent evaluations.

This problem of criterion contamination in confession-based cases is an important one. However, the study also included cases that used factual analysis to establish ground truth. In these cases the criterion that led to their inclusion did not make use of evaluation of the behaviors that were assessed by the evaluators. It is possible, however, that the suspects' behaviors that were evaluated (by the four evaluators) may have been related to the activities that were incorporated in the factual analysis review and therefore were contaminated in that indirect way. For that reason, there was some, albeit small, possibility of contamination. Thus, while there may have been contamination of both of the major criteria used in the study, it is important to note that the two quite different criteria did not produce significant differences in evaluators' outcomes.

Ground truth for two of the suspects included in this study was established not by confession or factual analysis but rather by incontrovertible evidence that was developed. In both instances, the suspects were found to have been "truthful" in their BAI. The decision in each instance was confirmed by the fact that the money the suspects were believed to have stolen from their employer, whose documentation erroneously indicated receipt, had actually been misdirected by the sending company to another location.

If it were possible to develop ground truth criteria in a large number of cases such as occurred in these two instances, the interpretation of findings would be less problematic.

Unfortunately, such cases do not occur frequently. Hence, while confessions—and perhaps factual analysis—are the most practical and dependable criteria for field-based research of the kind reported here, their limitations must be carefully considered. For this reason, these findings should be cautiously generalized to real-life settings.

In spite of the caveats to be considered, these findings support the observations of field practitioners with respect to the Behavior Analysis Interview. The BAI appears to be a procedure in which persons with a professional interest in sorting truth-tellers from liars, including, of course, almost all police investigators, would benefit from training. Further research is necessary to determine the effect of such training on the ability to elicit and to judge behavior and on the contribution that it may make to investigator's performance.

References

[1] Eck, J., *Solving Crimes: The Investigation of Burglary and Robbery*, Washington, D. C., Police Executive Research Forum, 1983.

[2] Ericson, R., *Making Crime: A Study of Detective Work*, Toronto, Canada, Butterworth and Company, 1981.

[3] Greenwood, P., Chaiken, J., and Petersilia, J., *The Criminal Investigation Process*, Lexington, MA, D. C. Heath and Company, 1977.

[4] Miyazawa, S., *Policing in Japan: A Study on Making Crime*. Albany, New York, State University of New York, 1992.

[5] Wilson, J., *The Investigators*, New York, Basic Books, Inc., 1978.

[6] Horvath, F., "Verbal and Nonverbal Clues to Truthfulness and Deception During Polygraph Examinations," *Journal of Police Science and Administration*, Vol. 1, 1973, pp. 138–152.

[7] Trovillo, P., "A History of Lie Detection," *Journal of Criminal Law, Criminology, and Police Science*, Vol. 29, 1939, pp. 848–881.

[8] Morris, D., *Manwatching—A Field Guide to Human Behavior*, New York, Harry Abrams Inc. Publishers, 1977.

[9] Nierenberg, G. and Calero, H., *How to Read a Person Like a Book*, New York, Pocket Books Publishers, 1971.

[10] Burgoon, J., Buller, D., and Woodall, W., *Nonverbal Communication: The Unspoken Dialogue*, New York, Harper & Row, 1989.

[11] Carpenter, R., "The Statistical Profile of Language Behavior with Machiavellian Intent or While Experiencing Caution and Avoiding Self-Incrimination," *Annals of the New York Academy of Sciences*, Vol. *606*, 1990, pp. 5–17.

[12] Ekman, P., *Telling Lies*, New York, W. W. Norton and Co., 1985.

[13] Ekman, P. and Friesen, W., "Nonverbal Leakage and Cues to Deception," *Psychiatry*, Vol. 32, 1969, pp. 88–106.

[14] Ekman, P., Friesen, W., and Scherer, K., "Body Movement and Voice Pitch in Deceptive Interactions," *Semiotica*, Vol. 16, 1976, pp. 23–27.

[15] Feldman, R., Devin-Sheehan, L., and Allen, V., "Nonverbal Cues as Indicators of Verbal Dissembling," *American Educational Research Journal*, Vol. 15, 1978, pp. 217–231.

[16] Knapp, M., Hart, R., and Dennis, H., "An Exploration of Deception As a Communication Construct," *Human Communication Research*, Vol. 5, 1974, pp. 270–285.

[17] Kraut, R., "Verbal and Nonverbal Cues in the Perception of Lying," *Journal of Personality and Social Psychology*, Vol. 36, 1978, pp. 380–391.

[18] Mehrabian, A., *Nonverbal Communication*, Chicago, Aldine Atherton, 1972.

[19] Streeter L., Krauss, R., Geller, V., Olson, C., and Apple, W., "Pitch Changes During Attempted Deception," *Journal of Personality and Social Psychology*, Vol. 35, 1977, pp. 345–350.

[20] Knapp, M. and Comandena, M., "Telling It Like It Isn't: A Review of Theory and Research on Deceptive Communication," *Human Communication Research*, Vol. 5, 1979, pp. 270–285.

[21] Ekman, P. and O'Sullivan, M., "Who Can Catch a Liar?," *American Psychologist.*, Sept. 1991, pp. 913–919.

[22] Reid, J. and Arther, R., "Behavior Symptoms of Lie Detector Subjects," *Journal of Criminal Law, Criminology and Police Science*, Vol. 44, 1953, pp. 104–108.

[23] Barland, G., "Detection of Deception in Criminal Suspects: A Field Validation Study, Unpublished dissertation, University of Utah, 1975.

[24] Horvath, F. and Jayne, B., "A Pilot Study of the Verbal and Nonverbal Behaviors of Criminal Suspects During Structure Interviews," Unpublished paper, U. S. Dept. of Defense Grant # 89-R-2323, 1990.
[25] Horvath, F., "The Effect of Selected Variables on the Interpretation of Polygraph Records," *Journal of Applied Psychology*, Vol. 62, 1977, pp. 127–136.
[26] Inbau, F., Reid, J., and Buckley, J., *Criminal Interrogation and Confessions*, 3rd ed., Baltimore, Williams and Wilkins, 1986.
[27] Reid, J. and Inbau, F., *Truth and Deception: The Polygraph (Lie Detector) Technique*, Baltimore, Williams and Wilkins, 1977.
[28] Jayne, B., "Factual Analysis in Employee Theft Investigations," unpublished manuscript prepared for *The Investigator Anthology*, Chicago, J. Reid & Associates, 1990.
[29] Kraut, R. and Poe, D., "On the Line: The Deception Judgments of Customs Inspectors and Laymen," *Journal of Personality and Social Psychology*, Vol. 39, 1980, pp. 784–798.

Address requests for reprints or additional information to
Frank Horvath, Ph.D.
School of Criminal Justice
512 Baker Hall
Michigan State University
East Lansing, MI 48824
(517) 355-2197
(517) 336-1787 (FAX)