



1976

Human Differences in the Ability to Differentiate Spoken Lies from Spoken Truths

Paul John Lavrakas
Loyola University Chicago

Recommended Citation

Lavrakas, Paul John, "Human Differences in the Ability to Differentiate Spoken Lies from Spoken Truths" (1976). *Dissertations*. Paper 1689.
http://ecommons.luc.edu/luc_diss/1689

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 1976 Paul John Lavrakas

HUMAN DIFFERENCES IN THE ABILITY
TO DIFFERENTIATE SPOKEN LIES
FROM SPOKEN TRUTHS

by

Paul J. Lavrakas

A Dissertation Submitted to the Faculty of the Graduate School
of Loyola University of Chicago in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

December

1976

ACKNOWLEDGMENTS

I thank my wife, Barbara J. Lavrakas, for the interpersonal support she has given to make my graduate work possible. I also thank my parents, John and Catherine Lavrakas, for the emotional support they have always provided me.

I thank Dr. R.A. Maier, Dr. E.J. Posavac, and Dr. Leonard Bickman for the opportunities they have provided for my growth as a research psychologist.

Finally, I thank the members of my dissertation committee, Dr. R.A. Maier (Chairman), Dr. E.J. Posavac, and Dr. T. Petzel.

VITA

The author, Paul John Lavrakas, is the son of John Lavrakas and Catherine (Gomatos) Lavrakas. He was born October 7, 1946 in Cambridge, Massachusetts.

His elementary and secondary education was obtained in the public schools of Birmingham, Michigan. He graduated from Birmingham High School in June, 1964.

In September, 1964, he entered Michigan State University and in June, 1968, received the degree of Bachelor of Arts with a major in social sciences.

In September, 1968, he began teaching in the Chicago Public Schools, where he remained employed until his resignation in September, 1972.

In September, 1972, he was granted a research assistantship in experimental psychology at Loyola University of Chicago. In February, 1975, he was awarded a Master of Arts degree in Psychology.

In June, 1976, he became Executive Director of Social Systems Research Institute, Evanston, Illinois.

His published works include: "Female preferences for male physiques" (1975); "Citizen crime reporting projects: a summary report" (1976); "Lying behavior and the evaluation of lies" (1976); and "A perspective on the recognition of other-race faces" (1976).

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
INTRODUCTION	1
Past Investigations of Human Lie Detection	3
Purpose of Present Study	17
Hypothesized Correlates of Judgmental Accuracy	21
Interpersonal Trust and Judgmental Accuracy	27
Summary	31
METHOD	33
Subjects	33
Making of Honest and Dishonest Audio-tape Stimuli	33
General Procedure	38
Dependent Variables	40
Subject Variables	42
RESULTS	48
Subject Reliability	48
Ratings of Honest/Dishonest Tape Stimuli	48
Correlation Analyses	58
Regression Analyses for Judgmental Accuracy	62
Discriminant Analysis for Accuracy Scores	69
DISCUSSION	74
Judgmental Accuracy	74
Judgments in Human Lie Detection and Cronbach's Measurement Components	80
Validity Issues	84
Possible Future Directions for Research in Perceiver-Differences in Human Lie Detection	93
Possible Applied Implications of Research in Perceiver-Differences in Human Lie Detection	94
REFERENCE NOTES	95
REFERENCES	96

	Page
APPENDIX A	102
APPENDIX B	104
APPENDIX C	106
APPENDIX D	109
APPENDIX E	111
APPENDIX F	113
APPENDIX G	115
APPENDIX H	118
APPENDIX I	120

LIST OF TABLES

Table	Page
1. Means and Standard Deviations of Subject Variables	49
2. Expected Versus Observed Frequencies of Accuracy Scores	52
3. Percentage of Accurate Judgments for the 20 Stimuli	54
4. Frequency Distributions of Scaled Judgments	57
5. Intercorrelation Matrix	59-61
6. Group One and Group Two Means for Subject Variables in Discriminant Analysis	71
7. Efficiency of Discriminant Function	72

LIST OF FIGURES

Figure	Page
1. Twenty Conditions for Generating Stimulus Lies and Truths	37
2. Rating Scale Used by Listeners to Judge Stimuli	41
3. Expected and Observed Frequency Distributions of Accuracy Scores	51
4. Scatterplot of Actual Accuracy vs. Predicted Accuracy	67
5. Two-Stage Judgment Model in Human Lie Detection	82

INTRODUCTION

The ancient commandment, "Thou shalt not bear false witness against thy neighbor", attests to the existence of lying (defined here as the conscious effort of a speaker to mislead a listener by making a false statement), and an awareness of the problems it causes, since earliest times. A recent report by the United States House of Representatives' Committee on Government Operations (House Report #94-795, 1976) reviewed some of the ways in which past societies have attempted to detect lies and verify truths:

At various times, and in different places, there evolved such tests as the ordeal of boiling water, the ordeal of the red-hot iron and the ordeal of the red-hot stove. In one such ordeal, the suspected wrong doer was expected to thrust his hand into a fire. If the hand was unsinged when removed the individual was declared innocent; if the hand was burned, that was positive proof of guilt. In other circumstances, truth or lack of truth was determined by the pattern assumed by a handful of tossed pebbles. A test used by the early Chinese required suspects to chew rice powder while being questioned. If the rice powder was dry when spit out, the man was condemned, on the premise that intention of guilt supposedly dried up his salivary glands. (p. 4)

As opposed to these various trials by ordeal (some of which do reflect a "primitive" understanding of psychology and physiology), the earliest known linking of emotional arousal with lie detection was made by Erasistratus (300 B.C.) who suggested that an increase in heartbeat-frequency was associated with deception. Despite this early observation, it appears that it was not until the late 1800's that a mechanized device was employed (by Cesare Lombroso, an Italian criminologist) in an attempt to detect deception (House Report #94-795, 1976; Wilhelm & Burns, 1954).

During the 1900's this interest in the efficacy of a machine to detect a physiological change, which was assumed to be associated with the act of lying, led to the development of sophisticated devices that monitor breathing pattern, blood pressure and pulse, and/or skin resistance (GSR). In addition, since the early 1960's two new devices have been developed, the Psychological Stress Evaluator and the Voice Analyzer, which attempt to detect deception by monitoring changes in various parameters of the human voice.

Despite these efforts to employ extremely sensitive technical devices for detecting deception, a United States Congressional Committee has concluded that "there is no lie detector, neither machine nor human" (p. 46). Thus a committee asserts that there exists, at present, no lie detector with proven validity (House Report #94-795, 1976).

Added to this controversy over the validity of mechanical lie detectors, Lykken (1975) has suggested that it may not be so much the machine which is or is not accurate as it is the human who operates the machine. Focusing specifically on the human facet, Maier and Lavrakas (1976) have recently observed that "the behavior of lying is something that most, if not all, humans have experienced both as a listener and as a liar.... Since the lying of others can cause problems for people, it is important to develop a sensitivity to lying so as to function efficiently in everyday life" (p. 581). In addition, the findings of Anderson (1966) highlight the importance that attempted-deception plays in person's lives. His results indicated that of 555 standardized personality-trait adjectives the ten least liked personality traits included liar (555), phoney (554), dishonest (551), untruthful (550), and

deceitful (546); while the ten most liked personality-traits include sincere (1), honest (2), loyal (4), truthful (5), and trustworthy (6).

Given these observations regarding the important role attempted-deception plays in interpersonal relations it can be suggested that humans have strong reason to learn to distinguish the behavior of a person engaged in lying from a person speaking the truth. This activity of trying to differentiate lying from truthfulness will be referred to as human lie detection, and, in a broad context, can refer to all human verbal interactions i.e., all verbal exchanges involve at least an implicit decision on the part of the listener of whether or not to believe the speaker.

From a behavioral perspective, human lie detection can be described as an interpersonal situation in which the speaker is the stimulus that the listener perceives. In the past decade research began within this stimulus-perceiver framework. This research has attempted to explore stimulus-properties, perceiver-properties and the interaction of stimulus and perceiver in human lie detection situations.

Past Investigations of Human Lie Detection.

In the late 1960's, N.R.F. Maier and his associates showed an interest in the systematic investigation of human lie detection because of the possible relevancy their findings might have within the context of personnel interviews. Maier (1966) devised a two-person role-playing method to generate honest interactions and dishonest interactions in order to investigate the question, "if one member of a pair makes honest versus dishonest statements does he behave differently and, if

so, can these differences be detected by the other member of a pair?" (p. 55). Maier's role-playing format involved one subject interacting in the role of a student with another subject playing the role of a professor. The content of the interaction dealt with the possibility that there had been a grading error on an exam, versus the possibility that the student had altered the exam after it had been graded. The role instructions for the professor were invariable; they explained that the student was concerned about the low grade he had received and felt that part of his answer had not been graded, as it had been written on the reverse side of a blue book page. It was explained in the professor's instruction that this answer could have been added after the blue book had been returned. It was also explained that the professor's assistant, who had graded the exam, felt certain that the answer on the reverse side had not been there originally.

The role instructions for the student-subject varied depending upon the condition to which he/she was assigned. The honest role-playing instructions informed the student-subject that the low grade on the exam appeared to be the result of the grader possibly not seeing a part that had been written on the reverse of a blue book page. As this part had been well written the honest student was told that he/she should reasonably have expected a better grade. The dishonest role-playing instructions informed the student-subject that the low grade on the exam would hurt his/her grade point and that, by adding a well written part to a reverse side of a blue book page, the professor could possibly be convinced that his assistant had missed this answer and therefore be willing to raise the grade on the exam.

Eighty-eight professor-student pairs interacted with the professor-subject unaware of the condition that his student-subject had been assigned. Following the interaction the professor-subject rated the student-subject on a five-point trust-distrust scale (with an "uncertain" midpoint). Of the 70 professor-subjects who made a trust or distrust decision (18 indicated that they were uncertain) 64.3% correctly judged the honesty or dishonesty of their student role-player. Maier presented the following discussion:

A clear finding in the experiment is that while the interviewers distinguish between honest and dishonest interviewees better than chance...preliminary explorations of cues used to judge the honesty of the interviewee found that the same cues led to opposite conclusions. Therefore the only conclusion justified from our present knowledge is that the interviewers get an impression and, on the basis of this, make better than chance judgments (p. 65).

In a follow-up study Maier and Jansen (1967) chose to investigate the "reliability of reasons used in making judgments of honesty and dishonesty" i.e., were there specific behaviors in the speaker (stimulus properties) that were systematically related to the decisions reached by the observer (perceiver)? The same student-professor role-playing format was used to generate four role-playing interactions which were observed by 57 subjects. These subject-observers made individual ratings regarding the honesty or dishonesty of the four student role players. In addition they indicated what behaviors of the student role players influenced their ratings. As found by Maier (1966), subjects as a group were significantly better-than-chance in accurately judging the honesty or dishonesty of the student role-players. Comparison of the accuracy of individual judges also showed that the judges differed in their judgmental accuracy i.e., accuracy ranged from no correct ratings to all correct

ratings. Despite the individual differences in judgmental accuracy the reasons given by subjects for making their judgments did not reliably relate to accuracy or inaccuracy. The only trend that was present was that accurate judges gave a "few" more reasons for their judgments than did inaccurate judges. Maier and Jansen concluded:

Do judges know the reason behind their judgments of honesty or dishonesty? Our answer must be a qualified 'no'. While very good and very poor judges used the same kinds of reason...the number of reasons presented relates more to certainty of judgment than to accuracy.... The sum total of the evidence indicates people make judgments with varying degrees of accuracy, but these judgments seem to be based upon impressions rather than logic (p. 150).

This study did not find properties inherent to the perceiver (as operationalized by the type of reasons given for the ratings) which were reliably related to accuracy in human lie detection.

Possibly because of the null finding of Maier and Jansen, Maier in his follow-up study (Maier & Thurber, 1968), changed directions and pursued an investigation of stimulus properties as they may relate to accuracy in human lie detection. Using the same student-professor role-playing format, Maier and Thurber varied the medium by which subjects received information about the honest and dishonest interactions: a) a Watchers group both watched and heard the interaction; b) a Listeners group heard tape recordings of the interaction; and c) a Readers group read transcripts of the interaction. The logic of this exploratory study was to "isolate visual, auditory, and verbal cues by comparing the accuracy of judgments under the three conditions,...as the speaker's personality, facial expressions, verbal intonation and mannerisms all may give cues which serve as either aids or distractors" (p. 24).

A total of 57 Watchers, 98 Listeners and 64 Readers judged two

honest and two dishonest interactions on a six-point trust-distrust scale. The results indicated that while all groups were significantly better-than-chance, the Listeners (average accuracy of 77%) and Readers (average accuracy of 77.3%) were significantly more accurate than Watchers (average accuracy of 58.3%). These results were interpreted as indicating that visual cues may serve as "distractors" and, as such, account for the inferior performance of the group that both saw and heard the honest/dishonest interactions. In drawing this conclusion Maier and Thurber were not explicit about whether the visual information "distracted" the Watchers by providing them more information to attend to, or "distracted" them because visual information is harder to judge than spoken information.

Despite this uncertainty about why the Watchers were less accurate, the authors drew attention to an interesting parallel between their findings and the "real world". Maier and Jansen stated:

Juries are composed of untrained observers who often must make judgments about the integrity of a witness. The witness is always present and, as these results suggest, may serve as a distractor. It is also interesting to note that the symbol of justice stands with...a blindfold over her eyes-she can only hear! (p. 30)

Independent of the research N.R.F. Maier and his coworkers were engaged in¹, P. Ekman and W.V. Friesen began research related to their hypothesis that there are body movements and facial expressions which a deceiver is not able to control and, therefore, emerge as leakage or deception cues. Specifically relevant to the Maier and Thurber "distraction-of-visual-information hypotheses" was Ekman and Friesen's

¹Personal communication from P. Ekman, July, 1976.

(1969) theoretical suggestion that: the face is equipped to lie the most ...and thus can be a very confusing source of information in deception" (p. 99). "The body, however, usually more truthfully reveals to the observer how the person actually feels (leakage) or the fact that something is amiss (deception cues)" (Ekman & Friesen, 1974, p. 289). As Maier and Thurber's Watchers were untrained observers they may well have focused on the face for information and, thereby, were distracted.

To experimentally pursue their hypothesis that the body was a greater source of deception cues than the face, Ekman and Friesen (1974) made a set of video-tapes which contained excerpts from both honest and dishonest interactions. These video-taped interactions, unlike the role-playing procedure used by Maier and his associates, conformed with Ekman and Friesen's (1969) four criterion dimensions of deception interaction. These criterion dimensions are: a) the saliency of the deception; b) the stakes for success; c) the balance of roles; and d) the extent of antagonism between deceived and deceived about the maintenance of deception.

Ekman and Friesen (1974) operationalized these criteria within the context of student nurses being interviewed about a series of films they were watching. Specifically, a group of student nurses were invited by the Dean of the Nursing school to individually participate in an experiment. At the time of the experiment it was explained to a student nurse that she would be asked to view two pleasant films and two stressful films (medical training films depicting amputation, and the treatment of severe burns). It was also explained to her that "a nurse must be able to deceive in certain situations e.g., when talking with the family of

a severely injured child, she must conceal her own worries or distress and convey positive affect to reassure the parents" (p. 290). Therefore the student nurse was led to believe that the deception she would be asked to engage in would be similar to deception that would be relevant to her "success" as a nurse.

While the student nurse was watching each film she was interviewed by an adult female who was unaware of the content of the specific segment being seen. Prior to the start of each segment the nurse was shown instructions informing her to be honest or deceptive in the subsequent interview. The interview itself dealt primarily with questions regarding the feelings the student nurse was experiencing while watching the film. While the interview film-session transpired video-tapes were made of the nurse's face and of her body by concealed cameras.

Following the filming of these interviews "composite" video-tapes were made from samples of the filmed interviews of 16 different student nurses for use in two subsequent judgmental tasks. Judgmental Task A required observers to rate one sample of non-verbal behavior from each nurse's nonverbal repertoire (p. 291). Judgmental Task B required observers to first view an example of a nurse's nonverbal behavior which was identified as honest (i.e., a familiarity example) and then rate an unidentified example of the same nurse.

Observers were assigned to experimental conditions in a 2 x 2 x 2 repeated-measurement ANOVA in which the variables were area (face or body), task (A or B), and condition (whether the behavior judged was actually honest or deceptive). The results indicated that only when observers had seen an example of the nurse's honest nonverbal behavior

were more accurate judgments made from the body versus the face. This was interpreted by the authors as partial support of their hypothesis. In addition to these results Ekman and Friesen had four experienced facial-analysts view the video-tapes of the nurse's faces with the familiarity example. "Each of these observers accurately judged both the honest and deceptive behavior of almost all the nurses, suggesting that the information is there in the face, if the viewer knows what to look for and how to interpret it? (p. 295).

Pursuing their interests in the stimulus properties that distinguish honest behavior from deceptive behavior, Ekman and Friesen (1976) investigated both visual and vocal behavior, using the set of video tapes from their previous study. These video tapes of the student nurses were analyzed for hand movement and voice pitch as follows:

Three types of hand movements were distinguished: illustrators, movements which are tied to speech rhythm and illustrate what is said; shrugs, ... in which the hands are rotated at the wrists to symbolically transmit the message of uncertainty or inability; and, adaptors, movements in which one hand makes contact with the other hand or other part of the body or the face. Pitch was measured by selecting two short speech samples from the [nurse's] answers and extracting fundamental frequency (pitch) of the voice by autocorrelation procedures using an on-line speech analysis computer system (p. 2).

Results of these analyses indicated that deceptive interactions contain significantly few illustrators, more shrugs, and an increase in voice pitch as compared with the honest interactions.

Ekman and Friesen also had four groups of subjects rate these interactions on 14 bi-polar scales e.g., trustworthy/untrustworthy. One group saw only the body, the second group saw only the face, the third only heard the interactions after they had been electronically filtered

to remove those frequencies above 400 Hz, and a fourth group only heard unaltered versions of the interactions. While the authors did not report results related to the accuracy of the various groups, they did find that there was no difference in the scale ratings of the group exposed to the filtered audio compared to the ratings of the group who were exposed to the unaltered voices. On the other hand, the group exposed to the face-videos made more positive ratings (e.g., more trustworthy) than did the group exposed to the body-videos. This last finding was interpreted as being in accordance with the previous findings of Ekman and Friesen (1974). Finally, the authors concluded, on the basis of the relationship between the voice pitch measurement and the observer ratings that the group which listened to the unaltered speech apparently were misled or confused by the speakers' control of their voices during deception.

Ekman and Friesen (1976) stated that forthcoming reports would describe the results of these findings in greater detail. In addition, they stated that they were presently working on research designed to replicate these findings on a new sample of "interactions" and, in general, to further investigate their interest in the stimulus properties in human lie detection.

As the problems of "Watergate" were causing national concern over the credibility of high officials, R.A. Maier and P.J. Lavrakas (1976) noted the general dearth of systematic research that had been done on human lie detection. As such, they began a series of studies to follow some of the leads of the earlier investigations, and to probe further into the factors affecting detection of lying.

Maier and Lavrakas chose to begin with a study designed to validate

the role-playing procedure used by N.R.F. Maier and his associates (Maier, 1965; Maier & Jansen, 1967; and Maier & Thurber, 1968). This approach was taken because of a methodological concern with the construct validity of the role-playing interactions i.e., did a person role-playing an honest or dishonest role validly portray the construct "truth" or "lie". In order to investigate this issue a lie detection apparatus (measuring GSR) was used to determine if individuals role-playing honest and dishonest roles could be physiologically differentiate. It was hypothesized that if the lie detector did differentiate, then people role-playing liars would presumably be similar to actual liars from a physiological-emotional standpoint.

In order to test this hypothesis Maier & Lavrakas had all subjects play both an honest and dishonest part, one as a student interacting with a professor (Experimenter One) and one as a motorist interacting with a policeman (also Experimenter One). Subjects were randomly assigned packets containing role-instructions so the experimenters did not know whether the subject was playing an honest or dishonest student or an honest or dishonest motorist. After Experimenter One and the sub-role-player had finished an interaction, Experimenter One asked the subject role-player some yes-no questions following a standard method used by polygraphers (Wilhelm and Burns, 1954, pp. 49-67). While the subject responded to the questions, Experimenter Two monitored the lie detector. Each subject's GSR responses in the honest and the dishonest role were later compared. The results indicated that subjects as a group were not consistently differentiated by the lie detector between their honest and dishonest roles. Either the lie detector

procedure was insensitive or subjects failed to react physiologically in the role-playing situations the same way that 'real' liars are assumed to react.

Using tape recordings of four role-playing interactions from this first study, Maier and Lavrakas next investigated whether listeners would correctly judge the role-players that the lie detector had correctly identified, and would fail to correctly identify the role-players which the lie detector missed. To this end, four taped role-playing interactions from the first study were chosen: two from a subject who had been correctly differentiated by the lie detector, and two from a subject who had not. A group of 56 listeners heard each tape recording and then rated it for honesty or dishonesty. It was found that subjects, as a group, correctly identified the two tapes of the role-player which the machine had been correct on. The group of listeners also correctly identified one of the two tapes of the role-player the machine had been wrong on. In addition, it was found that female listeners were marginally superior to males in their overall judgmental accuracy. The finding that subjects as a group were better-than-chance was consistent with those of N.R.F. Maier and his associates; but the null results of the first study placed doubt upon the construct validity of the role-playing technique. Thus it was uncertain whether it was lie/truth cues the listeners were basing decisions on, or whether it was more "difficult" to role-play a dishonest role and listeners were cueing in on this difficulty?

A final study reported by Maier and Lavrakas compared the judgment of honesty or dishonesty made by individuals with judgments made by a

group composed of these same individuals. It was reasoned that according to the diffusion-of-responsibility hypothesis (Bem, Wallach, and Kogan, 1965) a group would make more suspicious judgments than the average individual in that group i.e., an individual in a group bears less responsibility if another person is falsely accused of lying than does the same individual making his judgment alone.

To test this hypothesis Maier and Lavrakas chose a tape-recorded role-playing interaction which had earlier been found to elicit an approximately equal number of honest and dishonest judgments. Subjects first listened to the tape recording, and then rated the tape individually before meeting as a group to arrive at a group rating. The results showed that for all groups, the group rating was more suspicious than the average individual rating of the subjects comprising that group. This finding was interpreted as particularly interesting since many decisions concerning individuals are made by juries and committees which stress group consensus. As such, it appeared that while an individual qua individual may have been willing to give the speaker the benefit of the doubt in judging honesty, the same individual qua group member was willing to conform to the group consensus of a more suspicious judgment.

Besides the research programs of N.R.F. Maier and associates, Ekman and Friesen, and Maier and Lavrakas there appears to be few other studies directly investigating human lie detection. In one such study, Baskett and Freidell (1974) investigated the influence of linguistic and extra-linguistic variables on the attribution of lying. They found that if the target person responded either too quickly or too slowly the subjects judged his response as a lie more often than if the response

delay was of an intermediate duration. Unfortunately no results were reported which related the accuracy of the listeners' judgments to the target person's linguistic behavior.

In another study, Knapp and Hart (1974) required undergraduate Vietnam veterans to participate in both an honest and a dishonest interview in which they argued for or against increasing veteran's educational benefits. Analysis of the video-taped interviews showed that the veterans fidgeted (e.g., touching the face, moving the legs) and spoke more hesitatingly in the dishonest condition. Unfortunately these results appear somewhat equivocal as they have uncertain external and construct validity, i.e., were the veterans equally qualified to argue against the benefits, or were they merely arguing a position with which they had less personal familiarity?

A final study to be reviewed is that of Hocking (Note 1). He investigated the detection of deceptive communication from verbal, visual, and para-linguistic cues, by operationalizing the lie/truth construct in a manner similar to Ekman and Friesen (1974). In this instance, subjects (all of whom planned careers in law enforcement) were invited to participate in an experiment by the director of their School of Criminal Justice. Similar to Ekman and Friesen, Hocking motivated his subjects by informing them that the ability to lie successfully was positively related to success as law enforcement officers. After a set of video-taped stimuli of honest/dishonest persons was made, 923 "human lie detectors" participated under one of 28 separate conditions. Some of the various conditions included seeing a close-up of the head only, or of the body only, or both the head and body; another condition dealt

with seeing the video-tapes in either color or black-and-white; and another condition dealt with hearing the audio or not hearing the audio.

The major results of his study were:

- (a) Judgments based on the body-only were more accurate than judgments based on the head-only within the context of lies and truths of emotional content. This result did not hold when the lies/truths were factual in content. Hocking therefore suggested that the Ekman and Friesen (1974) finding of body-leakage may be limited to emotional situations.
- (b) Within the head-only condition, judgmental accuracy was higher for the group that saw the tapes in color versus the group who viewed the tapes in black and white. While Hocking did not explicitly interpret this result, he did suggest that it was not due to the subjects being more motivated in the color format.
- (c) For the lies and truths that contained factual content, the audio-only group had a greater accuracy than the visual-only group. On the other hand, for the lies and truths that contained emotional content there was no relationship between audio or visual information and accuracy.
- (d) Observers who both saw and heard the interactions were significantly more confident in their judgment than those who only saw the interaction. In addition, those observers who saw head-only were more confident than the observers who saw head-and-body or body-only.

The review of the literature indicates a relevant dearth of definitive research which directly investigates human lie detection. Most

studies have focused on delineating behaviors that differentiate between a person who is lying and a person who is telling the truth. This interest has been referred to in this dissertation as the investigation of stimulus properties of the interpersonal behavior of lying. In general there appears to be a consensus of findings that indicate "liars" do behave differently from "truthtellers", and that this difference is discriminable by other humans. While these findings are very much in accordance with the assumption underlying mechanized lie detection (i.e., a person who is lying will show some identifiable behavioral reaction), there is some question as to the external validity and construct validity of the findings. Devising experiments that require subjects to "lie" or to tell the "truth" may not validly represent the veracity dimension as it exists in the real world. In addition, it is uncertain to what other settings and other types of "lies and truths" these findings are generalizable.

Purpose of the Present Study

The review of past investigations of human lie detection indicates that subjects as a group, are able to judge veracity better-than-chance. While this may not be surprising from a commonsense standpoint, it is especially relevant given the present concern over the validity of mechanical lie detectors. If the use of mechanical lie detectors is discontinued as the Committee on Government Operations of the United States House of Representatives (House Report #74-795, 1976) has recommended "for all governmental agencies for all purposes" (p. 46), then decisions regarding veracity shall remain totally within the province of human decision makers. With this in mind, it is of interest to know more about

perceiver differences in human lie detection.

As has been mentioned, persons appear to differ in their ability to accurately judge veracity. Maier and Lavrakas (1976) also found that females were marginally better than males and suggested that this could be due to females traditional acculturation as "social-emotional specialists". Yet other than this finding of possible sex differences there appears to be very little known about personological differences as they may relate to accuracy in human lie detection. Thus the present work was primarily designed as an exploratory investigation of human differences in the ability to judge veracity.

In the absence of past findings directly related to perceiver differences in human lie detection, a review of past research on human judgmental accuracy in person perception, was undertaken. It was intended that this approach would identify a set of personological variables which would allow for the investigation of patterns of individual differences as they relate to judgmental accuracy in human lie detection.

Immediately relevant to this aforementioned concern, Cline and Richards (1960) investigated whether or not accuracy in person perception was a general trait. They devised an on-the-street interview procedure in which an interviewee was questioned regarding his personal values and self-concept. Each interview was filmed, and, following its completion, the interviewee completed a series of personality inventories. In the test situation, a group of subjects viewed the filmed interviews, and then made judgments regarding each interviewee's personality. Cline and Richards concluded that while their results identified two independent components of accuracy in person perception it was nevertheless meaning-

ful to regard the ability to perceive others accurately as a general trait.

Supporting Cline and Richard's conclusion, Taft (1955) had earlier published a complete review of findings related to correlates of the ability to judge others. Overall, Taft concluded, as Allport (1937, p. 512) had suggested that while a person's ability to accurately judge others is in part a function of the specific situation the judgment is made in, there is nonetheless reason to consider this ability more of a general trait than as a specific one. Taft summarized his literature review by suggesting that:

The following characteristics are fairly consistently found to be positively correlated with the ability to accurately judge personality characteristics of others: a) age in children; b) high intelligence and academic ability (with analytic judgment especially); c) specialization in the physical sciences; d) esthetic and dramatic interests; e) insight into ones peers on specific traits; f) good emotional adjustment and integration (analytic tasks only); and g) social skills (only with tests of ability to predict subjects' behavior). The ability to judge correlates negatively with judges social dependence and his 'psychasthenic' score on the MMPI. Characteristics showing fairly consistent lack of correlation are age (in adults), sex, and training in psychology. Some possible relationships on which more evidence is required...are number of older siblings [negatively correlated with accuracy], literary ability [positively correlated with accuracy] and being a clinical psychologist [negatively correlated with accuracy] (p. 20).

More recently, Hunt and Lin (1967) investigated subjects' accuracy in judging personal attributes after listening to tape recordings of stimulus persons. The authors concluded that "our findings do provide evidence of accurate judgment of personal attributes from speech, and also some evidence of consistency of individual performance across samples of speech expressive of quite dissimilar personalities" (p. 452). In addition, Hunt and Lin found no reliable relationships between judg-

mental accuracy and whether the judge was open- or closed-minded.

These findings serve as background for the conceptualization of the present study. Specifically, it was decided that the veracity construct would be operationalized by tape recordings of persons encouraging (or discouraging) others to (or from) participating in an "experiment". The method employed to generate these lies and truths led a subject to believe he/she was actually helping the experimenter by advocating a certain position. In addition, it was decided to present the lies and truths via audio-tapes so as to exclude visual cues. This decision followed Maier and Jansen's (1968) and, more recently, Krauss, Geller, and Olson's (Note 2) findings that listeners are significantly more accurate than watchers. Thus to increase the likelihood that subjects would show variation in their judging ability a listening task was chosen for this dissertation that prior research suggested should maximally discriminate among human lie detectors.

The work of Vernon and Cline (1960) and of Taft (1955) lends support to the plausibility of the general hypothesis that there are patterns of individual differences associated with judgmental accuracy in human lie detection. Based upon this reasoning, and upon a review of literature relating specific personological variables to general judgmental accuracy, the following set of variables was chosen to serve as predictor variables for judgmental accuracy in human lie detection:

- (a) Field dependence/independence;
- (b) General intelligence;
- (c) Social intelligence;
- (d) Extroversion;

- (e) Neuroticism;
- (f) Repression/Sensitization;
- (g) Locus of control;
- (h) Dogmatism;
- (i) Artistic interests; and
- (j) Demographic and background variables.

In addition, a new method employed was devised to provide real lies and truths which subject-perceivers would later judge for veracity. This method was devised with the intention of broadening the knowledge currently available regarding the external and construct validity of human lie detection investigations.

Hypothesized Correlates of Judgmental Accuracy

Field Dependence/Independence. Gary (1967) stated that, "in a field-dependent mode, perception is inferred to be strongly dominated by the overall organization of the field, whereas in a field-independent mode the parts of the field seem to be experienced as discrete from the organized background and can be organized separately" (p. 5190). Gary investigated the implications of this statement within the context of susceptibility to social influence. His results indicated no relationship between field dependence/independence and susceptibility. Nightengale (1967), on the other hand, reasoned (on the basis of early investigations) that field dependent persons are more oriented towards interpersonal cues than are field independent ones and therefore would show superior judgmental accuracy. In general, this study also produced null results.

These conflicting hypotheses are not unlike the ones encountered

while performing a recent investigation of field dependence/independence and other-race recognition (Lavrakas, Buri and Mayzner, in press). In this instance there were past findings to support both sides of the field dependence/independence hypothesis. For the present study it was reasoned that the listener's role in human lie detection is not unlike the visual task in traditional field dependence/independence measures i.e., one must make a decision about some specific aspects in the entire stimulus-field despite the presence of a "distracting" background. The parallel is that the listener in human lie detection must also make a judgment which may be based upon the evaluation of specific aspects of the speaker's behavior, which the listener may distinguish from the gestalt. It then follows that field independent individuals may be better able to distinguish specific cues on which to base their judgments of honesty or dishonesty. As such, it is hypothesized that field independent persons will demonstrate superior accuracy as listeners in human lie detection. This hypothesis is also in accordance with Taft's (1956) findings that field independence was marginally related to the ability to accurately judge others.

General Intelligence. Allport (1937), in reviewing earlier investigations of the relationship between judgmental accuracy in person perception and general intelligence, concluded that "understanding people is largely a matter of perceiving the relationship between past and present activity, between expressive behavior and inner traits, between cause and effect, and general intelligence is the ability to perceive just such relationships as these" (p. 514). In a similar vein, Levanthal (1957) suggested:

The predictions of an observer about behavior of another person are a product of the observer's hypothesis or conceptualization regarding a particular individual. The accuracy of his predictions depends upon the validity of the conceptualizations. His hypotheses are, in turn, a function of the information available about the person to be judged, and of the judge's typical way of categorizing or forming concepts about others (p. 176).

These statements of Allport and Levanthal suggest that general intelligence may be positively related to judgmental accuracy in human lie detection. This hypothesis is based upon the following reasoning. In deciding whether or not to believe a speaker, a listener must put a priority on the importance of the speaker's behavior cues within a specific interaction. The listener then may make decisions based upon his/her evaluation of the most important cues. It is therefore hypothesized that persons high in general intelligence will make such decisions with superior accuracy than those relatively lower in general intelligence.

Social Intelligence. Taft (1955) concluded that judging ability appeared to be a combination of general intelligence and social intelligence. It follows from this and from common sense that a person who is high in social intelligence, i.e., a person who can accurately assess the dynamics of interpersonal behavior, should also be capable of accurately assessing the gestalt of a verbal exchange. Therefore, it is hypothesized that persons high in social intelligence will be more accurate as perceivers in human lie detection than persons who are relatively lower in social intelligence.

Extroversion. Taft (1956) and Vingoe and Antonof (1968) reported results that extroversion was negatively correlated with the ability to accurately judge others. Taft suggested that a less extroverted person is more socially detached and thereby may be more objective in his/her

social judgment. This reasoning can be extended to suggest that an extroverted person is more likely to be actively engaged in social relations and, therefore, less likely to take time to "see what's going on". On the other hand, a more introverted person is less likely to be actively engaged in social relations and may, in his/her detachment, develop a clearer perspective of the underlying dynamics of the social situation he/she is observing. While this extension of reasoning is speculative it is none-the-less hypothesized that persons lower in extroversion (i.e., more introverted) will have superior judgmental accuracy in human lie detection than those persons relatively higher in extroversion.

Neuroticism. Taft (1955, 1956) and Vingoe and Antonof (1968) also found evidence to support their hypothesis that well-adjusted, emotionally stable persons show a superior ability to accurately judge others than do relatively neurotic, emotionally unstable persons. Taft (1955) reasoned that: "the well-adjusted person is less subject to projecting himself into others than the poorly adjusted person and therefore is able to judge better" (p. 13). In accordance with this reasoning it is hypothesized that stable, well-adjusted persons will show a superior judgmental accuracy in human lie detection than would a more neurotic person.

Repression/Sensitization. In a review article of repression/sensitization as a dimension of personality, Byrne (1964) described Repressors as individuals who have difficulty perceiving threatening material accurately due to an avoidance of anxiety arousing stimuli. Sensitizers, on the other hand, are individuals who "perceive threaten-

ing stimuli as accurately or more accurately than neutral stimuli" (p. 172). This may be due to their approach toward anxiety arousing stimuli. Within the context of human lie detection it is reasonable to suggest that the potential of being lied to can be threatening to the listener, i.e., the listener may receive false information, in addition to being duped by a liar. As such it is hypothesized that Repressors will react to the task of judging veracity with more avoidance behavior, which in turn will make them less sensitive to the speaker's cues. Sensitizers, on the other hand, should show more approach behavior to the speaker's cues, thereby being more accurate in judging veracity than repressors.

Locus-of-Control. McDonald (1973) in reviewing the locus-of-control literature concluded that "internals engage in more instrumental goal-directed activity, whereas externals more often manifest emotional non-goal-directed responses" (p. 171). Following from this it was reasoned that an internally controlled person will, as a listener, be more likely to focus on the relevant speaker cues in human lie detection. An externally controlled person, on the other hand, should be more likely to approach lie detection in a more emotional, less goal-directed manner. Thus, it is hypothesized that individuals with an internal locus-of-control will be more accurate in judging veracity than individuals with an external locus-of-control.

Dogmatism. Rokeach (1960) described the openness of one's belief system as "the extent to which a person can receive, evaluate, and act on relevant information received from the outside on its own intrinsic merits, unencumbered by irrelevant factors in the situation arising from

within the person or from the outside" (p. 57). This open/closed mindedness dimension corresponds to low-dogmatism and high-dogmatism, respectively. As was earlier mentioned, Hunt and Lin (1967) had found no relationship between open/closed mindedness and the ability to accurately judge people in general person perception. In addition, another study, in which the person perception task was operationalized differently from Hunt and Lin, found no relationship between accuracy and dogmatism (Sawatzky and Zingle, 1969). Despite these null results, it was reasoned that an open minded listener may be more open to the entire gestalt of the speaker in human lie detection. Thus it is hypothesized that open minded (i.e., less dogmatic) persons will more accurately judge veracity in human lie detection than will close-minded (i.e., more dogmatic) persons.

Artistic Interests. Taft (1955) concluded that the ability to judge others seems to be higher in those persons who have dramatic and artistic interests. Taft's conclusion concurred with the findings of Hastorf, Schneider and Poletka (1970) in their review of past findings linking artistic interests with judgmental accuracy. The assumption underlying this relationship appears to be that persons with high artistic interest are more likely to be "sensitive" people i.e., by having an appreciation for the arts and their subtleties, such a person may also have an general appreciation for the subtleties of human emotions. Following from this speculation, it is hypothesized that persons with high artistic interest will be more "sensitive" as listeners in human lie detection and will perform more accurately than persons with less artistic interest.

Demographic and Background Variables. From an exploratory and descriptive standpoint the inclusion of certain demographic and background variables as predictors is deemed appropriate. The effect of birth order and family size upon an individual has been a continual interest of psychology. In addition to such sibling information other basic demographic information, e.g., sex, age, years of schooling, etc., and family background information is hypothesized to possibly be related to judgmental accuracy in human lie detection. Another area of a person's background that is speculated as possibly being related to accuracy is a person's habits with regard to reading, television, and movie going. Finally it is hypothesized that a person's own attitudes toward religion and moral issues may relate to his/her accuracy in human lie detection.

In general these hypotheses were speculative (based upon my own conjecture). Therefore no statement is advanced to the direction (positive or negative correlation) of any relationships between those demographic and background variables and accuracy in human lie detection.

Interpersonal Trust and Judgmental Accuracy

The investigation of listeners' judgmental accuracy in human lie detection is also an indirect investigation of listeners' trust of the speakers. Interpersonal trust, as defined by Rotter (1971), is an individual's generalized expectancy that "the word, promise, verbal or written statement of another...can be relied upon" (p. 444); this corresponds to what Giffin (1967) refers to as the reliability dimension in interpersonal trust as it relates to the communication process. Given the obvious reality that humans do not differentiate lies from truths with the same ease and accuracy they differentiate white from

black or hot from cold, it is hypothesized that listeners will show a tendency to be more trusting in their average judgments than distrusting. This hypothesis follows in part from a consideration of the acculturation to presume innocence when in doubt, and from the conjecture that listeners in general would rather be duped by a liar than disbelieve a truth teller if, as operationalized in the present study, they have nothing to lose by being mistaken. In addition to this reasoning, a pilot study for the present dissertation, employing a similar methodology, found listeners (when confronted with an ambiguous² judgment task) showed a tendency to be more trusting than distrusting.

As individual difference measures were to be collected to investigate their relationship to judgmental accuracy, it was reasoned that some of these same measures may relate to the degree of trust a listener demonstrates in human lie detection. In addition, the aforementioned pilot study found a significant correlation between the degree of trust listeners showed and their degree of Machiavellianism and level of trustworthiness of others. To this end, the following constructs were chosen to serve as predictor variables for interpersonal trust:

- (a) Trustworthiness of Others;
- (b) Dogmatism;
- (c) Machiavellianism; and
- (d) Background and Demographic Variables.

²Specifically the ambiguity of the pilot study's judgment task resulted from listeners being told that the set of honest/dishonest tape-recordings they were to judge could contain any proportion of lies and/or truths.

Trustworthiness of Others. A person's general belief of the degree others are trustworthy has been investigated by Rotter (1966) and Wrightsman (1974), who both developed scales to measure this dimension. Their results indicate that people do show variance on this dimension. Wrightsman (1966) also reports results that indicate that persons who in general regard others as more trustworthy, reflect this disposition in their behavior. In the present study, it was reasoned that those persons who have a high level of trustworthiness of others, will reflect it in their judgments regarding veracity. Specifically, it is hypothesized that persons high in trustworthiness of others will be more trusting in their judgments than persons low in trustworthiness of others.

Dogmatism. Deutsch (1960) found that "in an ambiguous situation involving the choice of trusting or not...low authoritarians are more likely to be trusting...while high authoritarians are more likely to be suspicious" (p. 140). The notion of dogmatism was advanced by Rokeach (1960) as a suitable way to conceptualize general authoritarianism. As such it was reasoned that dogmatism (general authoritarianism) may relate to the general trust or distrust a listener showed in his/her judgment of veracity. Specifically, it is hypothesized that persons low in dogmatism will be more trusting in their judgments than persons high in dogmatism.

Machiavellianism. As previously mentioned, a pilot study for this dissertation indicated a significant relationship between Machiavellianism and the general level of trust in listeners' judgments. This exploratory hypothesis had been advanced by R.A. Maier (Note 3) on the following reasoning:

Machiavellianism (Christi, 1968) refers to the degree that a person feels others are manipulable in interpersonal situations. Christi had found that a person who regards others as highly manipulable will in turn be more likely to be manipulating in his/her personal behavior. It was therefore speculated that high Machiavellians may project themselves into their judgment of others, in a way as to make high Machiavellians more suspicious of others' true intentions.

Following from this, it is hypothesized that listeners who are low in Machiavellianism will be more trusting in their judgment of the honest and dishonest tape-recordings than will persons high in Machiavellianism.

Demographic and Background Variables. Rotter (1971) reviewed past research relating individual differences to behavioral trust or distrust shown toward others. He cited findings that youngest children, and agnostics and atheists are less trusting of others. Other than these results, most experimental research investigating behavioral trust has focused primarily on the situational determinants of trust/distrust (e.g., Wrightsman, 1966). Because of this relative dearth of past findings regarding the relationship between demographic-background variables and behavioral trust/distrust, no specific directional hypotheses are advanced for the present study.

Differential Trust. The aforementioned pilot study also indicated that listeners, as a group, judged one stimulus person in a manner inconsistent with their judgments of the other stimulus persons. An investigation of this indicated that the "deviant" stimulus person's speech had an accent which was markedly different from the other stimulus persons' speech. As the stimulus tape-recordings for the present study were to be generated at a university with a heterogeneous (racial-ethnic) undergraduate population, an a priori decision was made to investigate whether differential trust would be shown toward stimulus per-

sons with "deviant" accents. Specifically, it is hypothesized that the trust shown by listeners in the human lie detection task will be lower for the group of stimulus persons with "deviant" accents as compared with the ratings of the group of stimulus persons without "deviant" accents. This hypothesized construct is here after referred to as differential trust.

Summary

Specifically, the following hypotheses were advanced in this dissertation:

- (a) Listeners, as a group, will show better-than-chance accuracy in judging veracity.
- (b) Field independence will be positively related to judgmental accuracy.
- (c) High general intelligence will be positively related to judgmental accuracy.
- (d) High social intelligence will be positively related to judgmental accuracy.
- (e) Extroversion will be negatively related to judgmental accuracy.
- (f) Neuroticism will be negatively related to judgmental accuracy.
- (g) Sensitization will be positively related to judgmental accuracy.
- (h) An internal locus of control will be positively related to judgmental accuracy.
- (i) High dogmatism will be negatively related to judgmental accuracy.
- (j) High artistic interest will be positively related to judgmental accuracy.
- (k) The above mentioned variables will combine in a multiple regres-

sion procedure to account for more of the variance in judgmental accuracy than any one single variable.

- (l) Listeners, as a group, will make average judgments in a more trusting, than distrustful, direction.
- (m) Trustworthiness of others will be positively related to trust in judgments.
- (n) Low dogmatism will be positively related to trust in judgments.
- (o) High Machiavellianism will be negatively related to trust in judgments.
- (p) Listeners, as a group, will show differential trust in their judgment of the "deviant" stimulus person, i.e., they will be more suspicious of these "deviant" stimuli.

METHOD

Subjects

Participants included 51 females and 52 males ranging from 15 to 33 years old. These 103 subjects came from five groups: (a) 36 Loyola University of Chicago undergraduate volunteers who received credit in an introductory psychology course for participation; (b) 27 high school students from a history class in Birmingham, Michigan³; (c) 17 persons enrolled in a summer session undergraduate statistics course at Loyola University; (d) 18 persons enrolled in a summer session undergraduate statistics course at St. Xavier's College, Chicago, Illinois; and (e) 5 adult volunteers. While this sample was not randomly selected it does constitute a heterogenous group of primarily caucasian (90%), middle-class, urban individuals.

Making of Honest and Dishonest Audio-Tape Stimuli

Following a procedure that had been used successfully in a pilot study a set of 20 tape recordings were made in which a person gave a testimony that was either the truth or a lie. It was important, for purposes of construct validity, that these taped recordings be actual truths and actual lies as opposed to "role-playing" testimonials. The manner in which each tape was made was as follows:

1. A subject entered an experimental room and was informed by the experimenter that he/she would be participating in a pilot study for an

³I thank Mr. John Petrakis and Mr. John Schultz of E.W. Seaholm High School for providing access to these students.



experiment that might take place later in the semester. The subject was informed that it was the experimenter's intention that this pilot study would serve as a practice time for the experimenter, and that the subject would provide the experimenter with an indication of how undergraduates perform in the subsequent task. At this time the subject was engaged in either an "interesting" task (e.g., playing an electric tennis game on television) or a "dull" task (e.g., watching the second hand on a timer for five minutes).

2. Following the completion of the task, the experimenter asked the subject to fill out a rating sheet to assess the subject's attitude toward the task. The subject was told that this would help the experimenter get an impression of how undergraduates react towards this task. The rating sheet contained several semantic-differential scales. Of central interest was the scale anchored with the words "very interesting" and "very dull". Using this rating scale it was possible to determine whether a subject had found the task interesting or dull. This was of critical importance as it provided the experimenter with a means of ultimately determining the subject's veracity. This rating sheet is presented in Appendix A.

3. Following the rating of the task, the subject was presented with one of two written "cover stories" whereby he/she was asked to give an honest or dishonest testimony about the "dull" or "interesting" task. By reading the cover story, the subject was led to believe that his/her testimony would be played to high school students in order to encourage (or discourage) these high school students to (or from) participating in the actual experiment. Subjects who said the task was interesting

and who had found their task "interesting" were presumably telling the truth. Subjects who had found their task "interesting" and who had said their task was "dull" were presumably telling a lie. Subjects who had found their task dull and said that it was "interesting" were presumably telling a lie. Finally, subjects who found their task dull and who said it was dull were presumably telling the truth. The "cover story" instructions which subjects read are presented in Appendix B. In addition to these conditions, some subjects were asked to provide their testimony in their own words (i.e., "impromptu" tape stimuli) while others were asked merely to read the script provided at the end of the cover story (i.e., "read" tape stimuli).

4. When the subject indicated that he/she was ready to record the testimony the experimenter started the tape recorder and the subject proceeded to provide the testimony.

5. Following the making of the tape recording the subject was given a second rating sheet (see Appendix A) to assess the subject's attitudes toward making the tape. Of central importance was the scale anchored with the words "very honest" and "very dishonest". The rating the subject made on this scale presumably indicated whether the subject felt he/she had told the truth or had lied. In this way the experimenter was able to tell which tapes were truthful and which were lies.

6. Finally the subject was debriefed and it was explained that the experimenter was primarily interested in playing the tape recording to other persons to see if these others could determine if the subject had been honest or dishonest. Each subject was then asked for his/her permission to use the recorded testimony. All subjects gave this permission.

The above mentioned procedure was employed until the 20 taped-recorded stimuli shown in Figure 1 had been generated.

In order to make these 20-taped stimuli it required approximately 40 undergraduates to individually perform the above outlined procedure (1-6). This was due to the following circumstances. Initially, subjects were "randomly" assigned to a tape condition, but some subjects provided invalid tapes. This occurred when a subject's rating on the interesting/dull scale and/or the honest/dishonest scale was not clearly in one direction. As such, an a priori decision was made not to use the subject's tape, as it did not clearly constitute a valid truth or a valid lie. Because of this the tape generating procedure was continued with new subjects until all 20 tapes were validly made. There was no observable pattern to the conditions that required more than one subject to make a valid tape.

In summary, 20 tape recordings were made which constituted 10 truths and 10 lies. I regard these audio-tapes as validly representing the constructs "spoken truth" and "spoken lie". This statement is based on the fact that two rating scales were used to screen out those tapes which did not clearly conform with the validity criteria.

~~These~~ 20 taped testimonies were then sorted into two groups: the eight "improvised" tapes and the 12 "read-script" tapes. This grouping was performed on an a priori basis because it was assumed that if subjects listened to the 20 tapes together, they might show biases in rating the read-script tapes possibly thinking that these tapes were more likely to be lies. To avoid the possibility of this bias it was decided that the two groups of tapes would be played separately.

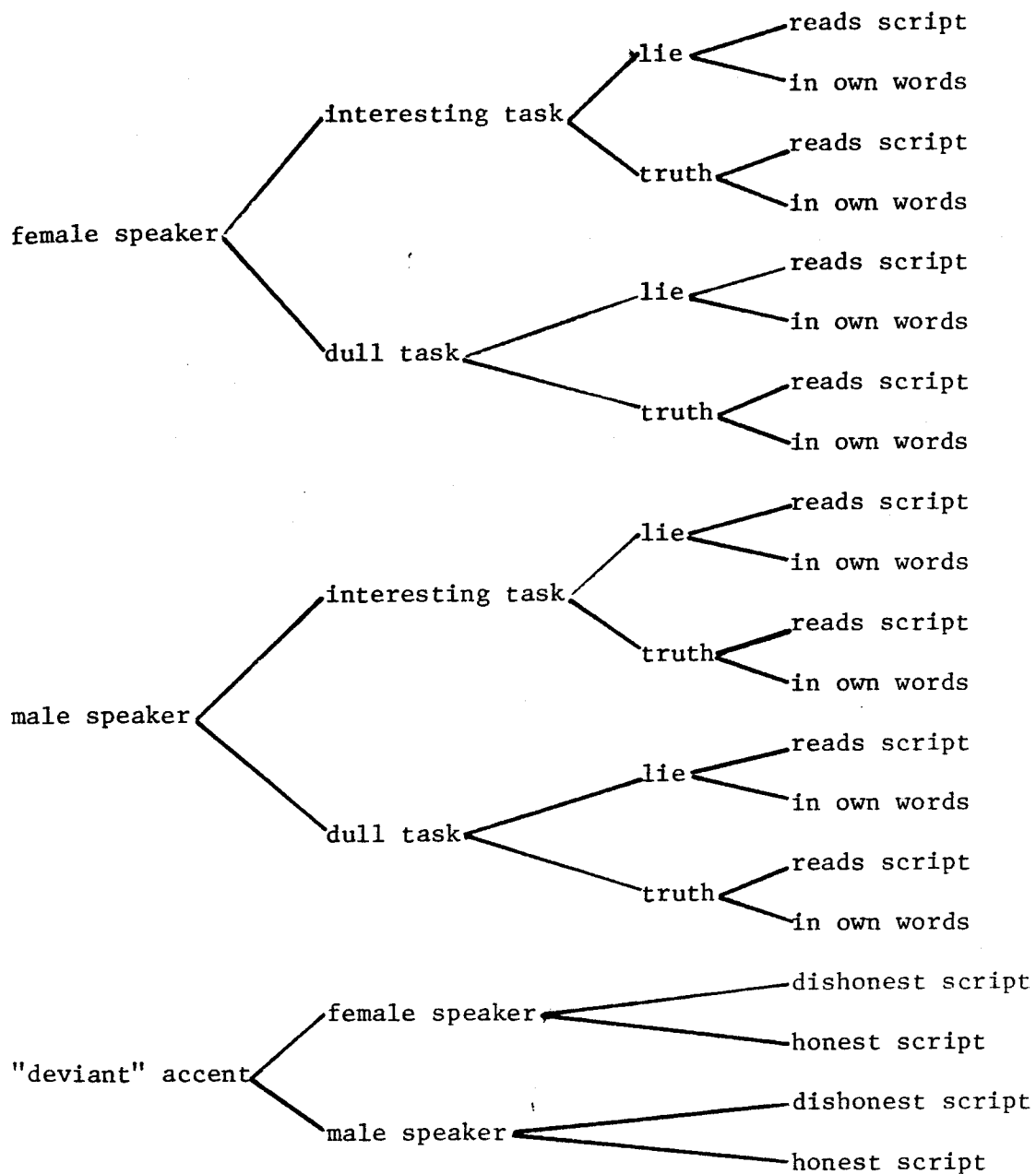


Figure 1. Twenty conditions for generating stimulus lies and truths.

The eight improvised tapes were randomly sequenced in several orders so that different listeners would hear them in different random orders. A similar random-sequencing was done for the 12 read-script tapes. These various sequences were then recorded with a Panasonic cassette recorder on a Memorex tape cassette.

General Procedure

Subjects, in groups of approximately ten, participated in the following manner (which was in part determined by time constraints, as subjects were unable to participate for more than two hours).

1. Instructions regarding the general nature of the experiment were provided verbally by the experimenter. These instructions explained that this experiment was an attempt to determine "who is a good lie detector". The experimenter also explained that this research was important because of its societal implications, and as such subjects were asked to sincerely participate to the best of their ability.
2. Subjects then took the Hidden Figures Test during a seven minute time period.
3. Subjects next listened to and rated the eight tape-recorded "improvised" testimonials. Prior to making their ratings, tape recorded instructions were played which acquainted the subjects with the task they were to perform. In addition, subjects were provided with an explanation of how the tape recordings had been made. The complete instructions are presented in Appendix C. The instructions also stated that the tape recordings had been randomly selected from a larger group of tape recordings so as to lead subjects to believe that the specific set of tapes that they would listen to might contain any proportion of lies

and truths. (In fact, the actual proportion was fifty/fifty.) Subjects listened to the taped voices one at a time and had approximately five seconds in which to make their own judgment before the next tape was played. The experimenter alerted the subjects that the next tape was about to begin approximately five seconds after the conclusion of the previous tape. The volume of the tape recorder was set loud enough so that all the subjects could hear the recordings adequately.

4. Subjects took the Cartoon Prediction Test during two four-minute timed intervals.

5. Subjects responded to Part 2 and Part 3 of the Cattell Culture-Fair Intelligence Test during timed intervals of two minutes and three minutes, respectively.

6. Subjects then heard tape recorded instructions explaining the next task, which was the rating of the "read-script" tape recordings. These instructions are similar to those heard prior to the rating of the improvised tapes, and are presented in Appendix D. After listening to these instructions, subjects listened to and rated the 12 "read-script" tapes in a manner similar to their rating of the eight "improvised" tapes.

7. Subjects, finally, were given a packet of personological inventories. They were instructed to complete it at their own leisure, but to return it as soon as possible. This packet included: (a) the Maudsley Personality Inventory; (b) a locus-of-control scale; (c) a dogmatism scale; (d) a trust worthiness of others scale; (e) a Machiavellianism scale; (f) an artistic interest inventory; (g) a demographic and background questionnaire; and (h) a modified repression/sensitiza-

tion scale. The packet in most instances was returned to the experimenter within one week. In all, subjects spent approximately 50 minutes participating in parts 1-6 and approximately one hour, at their leisure, responding to the packet of questionnaires.

Dependent Variables

The 20 tape recordings were rated using a scale (shown in Figure 2) similar to Maier and Jansen (1967). This scale provides each subject an opportunity to record his/her judgment of veracity, and his/her degree of confidence in the judgment.

Accuracy Score. Based on the actual scale ratings, each subject received an "accuracy score" which was computed by assigning a subject a "1" for each of the 20 stimulus tapes whose veracity was correctly judged. For example, if a tape were a lie and a subject marked the "lying" half of the scale the subject received a "1". On the other hand, if the tape were a truth and the subject marked a "lying" half of the scale, the subject received a "0". In this way an Accuracy Score was computed for each subject by summing his/her accuracy scores for each of the 20 judgments. This Accuracy Score could range from 0 (indicating a subject was incorrect on all 20 tapes) to 20 (indicating perfect accuracy).

Confidence Score. For every tape rating a subject had the option of whether he/she was "uncertain", "suspect", or "certain" in the judgment. By assigning a "1" to each uncertain rating, a "2" to each suspect rating, and a "3" to each certain rating, a "confidence score" was determined for every tape rating. By summing a subject's 20 confidence scores a Confidence Score was generated which reflects how confi-

- _____ I am certain this person is lying.
- _____ I suspect this person is lying.
- _____ I am uncertain but guess that this person is lying.
- _____ I am uncertain but guess this person is telling the truth.
- _____ I suspect this person is telling the truth.
- _____ I am certain this person is telling the truth.

Figure 2. Rating scale used by listeners to judge stimuli.

dent a subject was in his/her total ratings. This confidence score could range from 20 (indicating all uncertain ratings) to 60 (indicating all certain ratings).

Trust Score. In making their tape ratings subjects were led to believe that the set of tapes they were listening to could contain any proportion of lies or truths. Considering these directions subjects were left with an ambiguous task: that is, they had to rely totally upon their own judgment in order to determine whether they should use the truth or lie half of the rating scale. Thus, it was possible for subjects to differ in the amount of "trust" they displayed in their ratings, i.e., which part of the scale they used more often. Therefore, a "trust score" from "1" to "6" was assigned for each judgment a subject made. A "1" was assigned for "certain this person is lying", a "2" was assigned for "suspect this person is lying", ..., and a "6" was assigned for "certain this person is telling the truth". A Trust Score was then computed by summing a subject's "trust scores" for the 20-judgments. This Trust Score could range from 20 (indicating a total usage of the "I am certain this person is lying" response) to 120 (indicating a total usage of the "I am certain this person is telling the truth" response).

Subject Variables

During the experimental session and later at their own leisure subjects provided personological information to be used as independent variables for the subsequent analyses.

Field Dependence/Independence. Field dependence/independence was measured using the Hidden Figures Test (Thurstone and Jeffreys, 1965); this inventory has a split-half reliability of .95. Due to a time

restriction it was necessary to reduce the time period for administering this test to seven minutes as opposed to the 10 minutes intended for the full test. This time alteration limited the range of the scores and probably decreased the variance. Despite these numerical restrictions there is no reason to believe that the shortened time lessened the construct validity. The full test requires a subject to make 196 decisions, and is scored right minus wrong. For each decision a subject must first view a figure (i.e., some geometric design) and then indicate whether that figure is or is not contained in a separate drawing (i.e., a more detailed geometric design). High scores on this inventory represent field independence and low scores represent field dependence.

Social Intelligence. The Cartoon Prediction Test (O'Sullivan and Guilford, 1966) was used to measure social intelligence (split-half reliability of .70). This test is made up of two 15-item parts, each of which a subject has four minutes to complete. Each item requires a subject to first view a stimulus cartoon picture then choose one of three alternative pictures which would most likely happen next. Scores on this test can range from 0 to 30, with high scores indicating high social intelligence.

General Intelligence. A general intelligence score for each subject was computed by summing the scores on Part 2 and Part 3 from Form B, Scale 2, of the Cattell Culture-Fair I.Q. Test (Cattell and Cattell, 1960); the entire test has split-half reliability of .91. While the complete test has four parts only two of the four parts were used because of time constraints. In addition, the time subjects were permitted to work on the two tests was restricted to three and two minutes,

respectively, as compared with the four and three minutes recommended in the test manual. Part 2 contains 14 items each of which requires a subject to determine which of five patterns do not fit with the other four. Part 3 consists of 12 items, each of which requires a subject to determine which one of five alternative patterns best completes a large stimulus pattern. The total score for these two parts can range from 0 to 26 with high scores representing high general intelligence.

Extroversion and Neuroticism. The Maudsley Personality Inventory (Eysenck, 1962) was used to measure introversion/extroversion and neuroticism. The inventory is made up of two intermingled scales of 24 items each. For each item a subject read a statement and then indicated whether the statement was true, uncertain, or false as a description of himself/herself. Possible scores on both scales can range from 0 to 48. High scores on the extroversion/introversion scale represent extreme extroversion, and high scores on the neuroticism scale represent high neuroticism. A subject responds to the MPI at his/her own pace.

Locus-of-Control. Locus-of-Control was measured by Rotter's (1966) I.E. Inventory; this has a test-retest reliability of .72. The test consist of 29 items, eight of which are fillers. For each item a subject selects one of two statements that he/she believes to be more true. The test provides scores which range from 0 to 21; low scores identify persons with an internal locus-of-control and high scores identify persons with an external locus-of-control.

Machiavellianism. The Mach IV inventory (Christi, 1968) measures the degree to which an individual feels others are manipulable in interpersonable interactions (split-half reliability of .79). This scale

consists of 20 likert-format items. Scores can range from 20 to 120 with high scores representing persons high in Machiavellianism.

Dogmatism. The Short Dogmatism Scale (Schultz, 1962) was used to measure dogmatism (reproduction co-efficient of .83). It consists of 10 likert-format items which were selected from Rokeach's D-scale (1960) for best meeting the criteria of unidimensionality, item consistency, and reproducibility. Scores can range from 10 to 60, with high scores representing persons high in dogmatism.

Trustworthiness of Others. This was measured by the Trustworthiness Scale from the Philosophy of Human Nature inventory (Wrightsmann, 1964). The Trustworthiness scale measures the extent to which people are seen as moral, honest, and reliable (both split-half and test-retest reliability of .74). It is made up of seven positively and seven negatively worded likert-format items. Scores can range from 14 to 84 with high scores indicating a person who believes that others are highly trustworthy.

Artistic Interest. The degree to which a person has an interest in "art" was measured by an improvised scale which I constructed (see Appendix E). This was necessary because other available instruments required corporate-scoring forms which I did not choose to purchase. By using results presented in the Strong Vocational Interest Blank test manual (1968) it was possible to identify 10 occupation types which show a strong positive correlation with artistic interests, and 9 occupational types which show strong negative correlations with artistic interests. These occupational types were then presented in written format to subjects who were asked to indicate whether they would "like", "be indif-

ferent to", or "dislike" working in a particular occupation. A score was then computed for each subject by assigning a "2" for each "like" response, and a "1" for each "indifferent" response to the occupational types which were positively correlated with artistic interests, and by assigning a "2" for each "dislike" response and a "1" for each "indifferent" response for the occupational types which were negatively correlated with artistic interests. Following from this, scores could range from 0 to 38. As this was the first time the scale was used; its reliability is unknown.

Repression/Sensitization. The personality dimension of repression/sensitization was to be measured by a shortened version of Byrne's (1961) scale for this construct. The full scale consists of 127 items from the MMPI and has test-retest reliability of .82. Due to time constraints, a random set of 20 of these items was chosen for use. Subjects were presented these items in written format and were asked to indicate whether the statements were true or false for them. Long after these data were collected it was discovered that the version of the MMPI (1970) from which I originally sampled the 20 items, was not numbered in a manner consistent with Byrne's items numbering (taken from the original 1943 MMPI). As such, only 6 of the 20 items I employed actually came from the R/S scale. Due to this mistake no valid R/S scores were available for use in the analyses, and thus the R/S hypothesis is not testable.

Demographic and Background Questionnaire. A personal history and miscellaneous attitudes questionnaire was constructed in order to gather this type of information. In general the questionnaire consisted of

descriptive items (e.g., age, sex, number of siblings), personal habits and experience items (e.g., number of hours of TV watched per day) and some miscellaneous religious and moral attitudes items. This inventory is presented in Appendix F.

RESULTS

Subject Reliability

Inspection of frequency tables for the dependent and independent variables, led to a decision to drop three subjects from the analyses. One subject showed no variance in his judgments of the tape recordings, which suggests he deliberately made an insincere effort. An acquaintance of mine, who served as a subject, scored in a deviantly high manner on some of the independent measures, leading me to believe that he may have taken and/or scored these inventories before. A third subject to be dropped was a student of mine. This subject, due to partial paralysis, is not able to perform on timed tasks in a manner that validly reflects her ability. Due to this it was felt best to drop her from the final sample. Deletion of these three subjects left a sample of 50 females and 50 males for the subsequent analyses. All results described in this report will therefore be based on a sample size of 100.

Description of Subject Variables Data Set

Table 1 presents the means and standard deviations of the various personological variables. In general there is adequate heterogeneity of scores on all of these variables.

Ratings of Honest/Dishonest Tape Stimuli

Accuracy Scores. An Accuracy Score was calculated for each subject representing the total number of tapes he/she correctly judged. This score could range from 0 to 20. (The rationale for combining the 20 ratings into one Accuracy Score is presented in Appendix G.) The

Table 1

Means and Standard Deviations of Subject Variables

Variable	Mean	SD
Field Dependence/Independence	68.14	16.57
Social Intelligence	23.44	4.31
General Intelligence	15.57	2.89
Extroversion	27.95	9.43
Neuroticism	26.36	9.37
Locus of Control	10.95	4.36
Dogmatism	34.32	6.11
Artistic Interests	21.78	5.53
Trustworthiness of Others	49.13	9.62
Machiavellianism	63.71	11.12
Age in Years	20.16	4.24
How Religious	5.43	2.49
Degree of Personal Moral Code	7.09	1.88
Average # Hours Read per Day	2.56	1.80
Average # Hours TV per Day	1.77	1.46
Average # Movies per Month	1.50	1.05
# Shakespearean Plays Read	4.19	3.81
# Years of Education	13.36	2.08
# Siblings	2.76	2.35

observed sample mean of the Accuracy Score distribution is 10.89 with a standard deviation of 2.39. This mean can be subdivided into the following submeans: (a) a mean of 4.08 (sd=1.285) correct for the ratings of the eight non-accented "read" testimonies; (b) a mean of 4.480 (sd=1.418) correct for the ratings of the eight non-accented "improvised" testimonies; and (c) a mean of 2.330 (sd=.933) correct for the rating of the four accented "read" testimonies.

As the probability that a person will be correct-by-chance on any one judgment is .5, the expected value of the Accuracy Score on the basis of chance would be 10 correct or $E(X) = 10$. Similarly the expected value of the mean of the distribution of Accuracy Scores would also be 10 or $E(M) = 10$. In order to compare the observed mean (10.89) with the expected mean (10.00), a one sample t -test was performed. This value ($t(99) = 3.724, p < .001$) is high significant and supports the a priori hypothesis that people, in general, have a better-than-chance ability to listen to a tape recording and accurately judge a speaker's veracity.

Comparison of the observed frequency distribution of Accuracy Scores with the expected (on the basis of chance) frequency distribution is shown in Figure 3 and Table 2. Inspection of this figure and table supports the conclusion that peoples' ability to accurately judge veracity is better-than-chance ($\chi^2(20) = 33.5, p < .03$). This is illustrated in Figure 3 by the general displacement of the observed Accuracy Score distribution toward the high accuracy end of the abscissa. In addition, the observed frequency distribution of Accuracy Scores is negatively

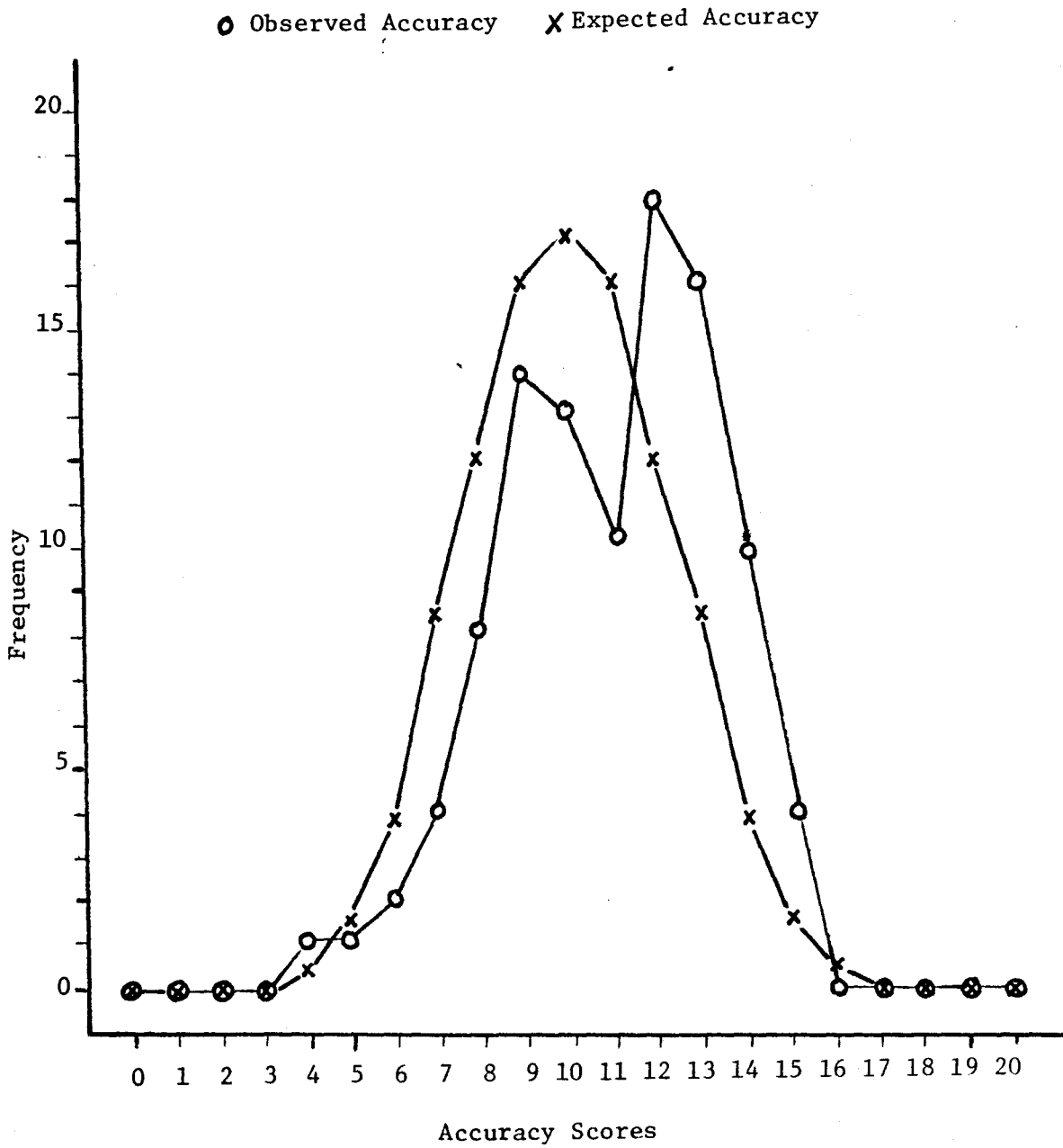


Figure 3. Expected and observed frequency distributions of Accuracy Scores.

Expected versus Observed Frequencies of Accuracy Scores

Accuracy Score	Expected f	Observed f
0	0.0	0
1	0.0	0
2	0.0	0
3	0.0	0
4	0.5	1
5	1.5	1
6	3.7	2
7	7.4	4
8	12.0	8
9	16.0	14
10	17.6	13
11	16.0	10
12	12.0	18
13	7.4	16
14	3.7	9
15	1.5	4
16	0.5	0
17	0.0	0
18	0.0	0
19	0.0	0
20	0.0	0
	<u>N=100.0</u>	<u>N=100</u>

Note. χ^2 (20) = 33.50, $p < .03$.

skewed⁴. Figure 3, then, demonstrates that it is not merely the central tendency of the distribution that is above the expected chance value, but the entire distribution is, in general, more accurate than chance.

These results represent significant findings, yet it is important to put them into the perspective of "meaningfulness". While it has been found that people are significantly better at differentiating honesty from dishonesty than chance-expectation, it is important to note the size of this significant difference. Table 3 shows the percentage of accurate judgments for each of the 20 taped stimuli; these percentages range from 32% correct to 77% correct with an overall mean accuracy of 54.45% correct. This observed mean percentage exceeds the expected percentage (50%) by only 4.45%. This difference is obviously not a large one. Therefore it is necessary to explicitly state that while the subjects, as a group, showed a significantly more accurate-than-chance ability to make their judgments, their average percentage correct is not impressive in absolute size.

Trust Scores. A Trust Score was computed for each subject by summing his/her rating on each of the 20 six-point judgment scales. This was done irrespective of the accuracy of the judgment. This Trust Score, therefore, reflects how "trusting" a subject was of the entire stimulus set. While the Trust Scores could range from 20 to 120, the observed range of the Trust Scores was 56-90. This observed distribution has a mean of 73.50 and a standard deviation of 7.386.

⁴Following a test of skewness (Snedecor and Cochran, 1967, p. 86) it was determined that the observed frequency distribution of Accuracy Scores is significantly skewed to the left (skewness statistic = $-.424$, one-tailed $p < .05$).

Table 3

Percentage of Accurate Judgments for the 20 Stimuli

Stimulus Tape	Description	Percentage Correct
1	male-read-lie	58%
2	female read-lie	32%
3	male-read-lie	37%
4	female-read-lie	43%
5	male-read-truth	45%
6	female-read-truth	67%
7	male-read-truth	61%
8	female-read-truth	65%
9	male-improvise-lie	36%
10	female-improvise-lie	59%
11	male-improvise-lie	59%
12	female-improvise-lie	51%
13	male-improvise-truth	41%
14	female-improvise-truth	63%
15	male-improvise-truth	77%
16	female-improvise-truth	62%
17	accent-male-lie	49%
18	accent-female-lie	63%
19	accent-male-truth	58%
20	accent-female-truth	63%

Note. Average percentage of correct judgments is 54.45%.

The expected value of this distribution can be looked at from three perspectives. First, if a subject were randomly responding to the tape ratings, it would be expected that his/her Trust Score would equal 70, i.e., an average of 3.5 on each of the 20 scale ratings. Secondly, if a subject rated all of the tapes accurately and with "certainty" he/she would also receive a score of 70, i.e., a "1" for the 10 lie tapes and a "6" for the 10 true tapes. Third, if a subject were uncertain as to what proportion of tapes were lies and what proportions were truths he/she could take a 50/50 chance strategy. Following this strategy, a subject would also be expected to score 70. These then are three lines of reasoning which lead to the same "expected value". The observed Trust Score mean of 73.5 deviates from the value of 70 in a "trusting" direction. By this, it is meant that subjects were more trusting than distrusting in their judgments. A one sample t -test shows this difference to be significant ($t(99) = 4.739, p < .001$).

In attempting to explain the significant difference it appears most likely that subjects, when faced with an ambiguous and difficult task, had a tendency to show inconsistent accuracy in their ratings and, in general, showed a slight, but significant tendency to be more trusting than distrusting in their total set of judgments.

Confidence Scores. A Confidence Score was computed for each subject, which indicated the overall degree of "certainty" a subject showed in his/her judgments. This score could range from 20-60. Had a subject shown no consistency in the degree of certainty of his/her judgments the expected value of this Confidence Score would be 40. In the same way, if the subjects as a group, showed no similarity in the degree of cer-

tainty the expected mean of the observed Confidence Score distribution would have also been 40. This value would represent the average rating on a given tape being made with a medium degree of confidence. The scores in the observed distribution ranged from 20-55 with a mean of 36.49 and a standard deviation of 6.74. A 99% confidence interval shows that one can be 99% certain that the mean of the Confident Scores falls within the 34.75-38.23 interval. As such it appears that while subjects as a group showed slightly less than medium certainty in their judgments, they did not feel they were merely guessing. This finding is also reflected in Table 4, which shows that of the total 2000 judgments, 17.20% were "certain", 48.05% were "suspect", and 34.75 were uncertain.

Differential Trust. By comparing the average trust score for the 16 non-accented tapes with the average trust score for the four accented tapes it is possible to compute a difference score which shall be referred to as "Differential Trust". Subtracting a subject's mean rating of the non-accented tapes from the subject's mean rating of the accented tapes provides a negative difference for a subject who was more trusting of the non-accented tapes, and a positive difference for a subject who was more trusting of the accented tapes. If subjects, as a group, had shown no consistent pattern of Differential Trust, the mean of these scores would have been zero. The actual mean, $-.2437$ ($sd=.7809$) was tested for deviation from the expected value using a one sample t -test. As a group, subjects were less trusting of the accented tapes than they were of the non-accented tapes ($t(99) = 3.121, p < .001$). This result is in the hypothesized direction. (In addition, Appendix H presents results relevant to the reliability of the Accent/Non-accent distinction.)

Frequency Distributions of Scaled Judgments

Stimulus Tape	Certain Lie	Suspect Lie	Guess Lie	Guess Truth	Suspect Truth	Certain Truth
1 (lie)	6	28	24	23	17	2
2 (lie)	3	11	18	23	34	11
3 (lie)	7	13	17	23	31	9
4 (lie)	8	20	15	19	34	4
5 (truth)	10	25	20	19	17	9
6 (truth)	4	13	16	7	37	23
7 (truth)	1	15	23	21	38	2
8 (truth)	4	15	16	16	33	16
9 (lie)	7	16	13	19	35	10
10 (lie)	9	30	20	14	18	9
11 (lie)	18	27	14	15	21	5
12 (lie)	10	26	15	16	25	8
13 (truth)	8	36	15	12	19	10
14 (truth)	9	22	6	14	31	18
15 (truth)	7	8	8	13	39	25
16 (truth)	9	19	10	14	31	17
17 (lie)	7	25	17	23	22	6
18 (lie)	12	27	24	20	13	4
19 (truth)	3	21	18	27	25	6
20 (truth)	<u>5</u>	<u>16</u>	<u>16</u>	<u>32</u>	<u>28</u>	<u>3</u>
Totals	147	413	325	370	548	197

Correlation Analyses

A set of Pearson product moment correlations were computed to investigate the bivariate relationships among the variables. These correlations are displayed in Table 5. Inspection of these results provides comment on the validity of the hypothesized relationships between the criterion variables (Accuracy Score and Trust Score) and the subject-predictor variables.

Subjects' ability to judge veracity, as reflected by the Accuracy Score, is found to correlate significantly with field dependence/independence ($r = .188$, one-tailed $p < .03$), and social intelligence ($r = .227$, one-tailed $p < .01$). These results indicate that field independence and high social intelligence are, individually, positively correlated with the ability to correctly judge veracity. In addition to these hypothesized relationships other personological variables are significantly related to Accuracy Scores:

- (a) The more siblings a person has the less accurate he/she is ($r = -.203$, two-tailed $p < .05$).
- (b) The more likely a person has a personal moral code which helps him/her make everyday decisions, the more accurate he/she is ($r = .202$, two-tailed $p < .05$).
- (c) The greater the average number of hours of television a person watches per day, the less accurate he/she is ($r = .243$, two-tailed $p < .05$).
- (d) The more Shakespearean plays a person has read, the more accurate the person is ($r = .203$, two-tailed $p < .05$).
- (e) The more confident a person is in his/her ratings, the more

Table 5

Intercorrelation Matrix

	AS	TS	CS	FDI	SI	GI	LOC
Accuracy Score (AS)							
Trust Score (TS)	.111						
Confidence Score (CS)	*.208	-.093					
Field D/I (FDI)	*.188	-.021	.072				
Social Intelligence (SI)	**.227	-.006	-.014	.103			
General Intelligence (GI)	.141	.044	.008	***.370	**.252		
Locus of Control (LOC)	-.124	.100	-.095	-.084	.023	-.070	
Neuroticism (NEU)	-.073	.017	-.058	*.192	-.104	-.055	**.226
Extraversion (EXT)	.044	-.149	.134	.046	.079	.076	-.107
Machiavellianism (MAC)	.060	.092	**.214	**.199	.102	-.107	***.294
Dogmatism (DOG)	.068	.042	**.207	.085	.159	*.171	.159
Trustworthiness (TRU)	.024	*.190	-.075	.063	-.044	-.026	-.018
Artistic Interests (AI)	-.030	-.001	-.028	.028	-.057	-.037	-.055
Sex (SEX)	-.080	.005	.088	.000	.014	.142	.048
Age (AGE)	-.017	.042	-.156	-.059	-.101	***.269	-.146
# Siblings (SIB)	-.203	*.168	-.040	.055	-.035	.096	-.077
Moral Code (MC)	**.202	.045	-.051	.162	-.016	-.002	-.093
Religious (REL)	.001	.044	-.138	-.103	.079	-.027	.048
Hours Read (RED)	.164	.133	-.060	.022	-.075	-.103	.026
Hours TV (TV)	***.243	-.042	.029	-.118	.119	-.029	.092
# Plays (PLA)	**.203	-.047	-.126	***.254	.147	.134	***.221
# Movies (MOV)	.147	.145	.055	***.288	-.025	.005	-.085

Note. Correlations are shown without decimal point.

*** $p < .01$, ** $p < .05$, * $p < .10$.

Table 5

Intercorrelation Matrix

	NEU	EXT	MAC	DOG	TRU	AI	SEX
Accuracy Score (AS)							
Trust Score (TS)							
Confidence Score (CS)							
Field D/I (FDI)							
Social Intelligence (SI)							
General Intelligence (GI)							
Locus of Control (LOC)							
Neuroticism (NEU)							
Extraversion (EXT)	-.211						
Machiavellianism (MAC)	.142	-.016					
Dogmatism (DOG)	.319	-.049	.156				
Trustworthiness (TRU)	-.171	.071	-.504	-.154			
Artistic Interests (AI)	-.230	.066	-.265	-.008	.126		
Sex (SEX)	.107	-.071	.212	.135	-.099	-.396	
Age (AGE)	-.101	.131	-.113	-.265	.045	.327	-.299
# Siblings (SIB)	-.016	.001	-.154	-.160	.068	-.091	.026
Moral Code (MC)	-.083	-.106	-.213	.032	.150	.152	-.251
Religious (REL)	.049	-.232	.130	-.080	-.352	-.101	-.057
Hours Read (RED)	.092	.013	-.036	.155	.082	.283	-.212
Hours TV (TV)	.277	.007	.019	.133	-.004	-.133	.117
# Plays (PLA)	-.051	.069	-.201	.101	-.002	.250	-.172
# Movies (MOV)	-.099	.134	-.057	.030	.063	.084	-.038

Table 5

Intercorrelation Matrix

	AGE	SIB	MC	REL	RED	TV	PLA
Accuracy Score (AS)							
Trust Score (TS)							
Confidence Score (CS)							
Field D/I (FDI)							
Social Intelligence (SI)							
General Intelligence (GI)							
Locus of Control (LOC)							
Neuroticism (NEU)							
Extraversion (EXT)							
Machiavellianism (MAC)							
Dogmatism (DOG)							
Trustworthiness (TRU)							
Artistic Interests (AI)							
Sex (SEX)							
Age (AGE)							
# Siblings (SIB)		-061					
Moral Code (MC)		.301 ***	-107				
Religious (REL)		.153	.002	-117			
Hours Read (RED)		.216 **	-.243 **	.309 ***	-103		
Hours TV (TV)		-.187 *	-.016	-.147	-.120	-.181 *	
# Plays (PLA)		.250 **	-.091	.243 **	.021	.279 ***	-.307 ***
# Movies (MOV)		.059	-.156	.151	.035	.078	-.056 156

accurate the person is ($r = .208$, two-tailed $p < .05$).

Regarding the level of trust listeners showed in the tape ratings (as operationalized by the Trust Score), few significant relationships were found. In fact, of the hypothesized relationships, only Trustworthiness of Others was found to be significantly related to Trust Scores ($r = .190$, one-tailed $p < .03$). While this correlation is significant its absolute size is surprisingly small. This statement is made in light of the fact that the Trustworthiness variable is suppose to reflect "how trustworthy one regards others" and the Trust Score presumably reflects how "trusting" a subject was in his tape ratings⁵. The only other variable that is marginally related to Trust Scores is the number of siblings a person has; in this case, persons with more siblings were less trusting of the stimulus tapes than were subjects with fewer siblings ($r = -.168$, two-tailed $p < .10$).

Regression Analyses of Judgmental Accuracy

In order to determine the best combination of subject variables for prediction of Accuracy Scores, a stepwise multiple regression analysis was performed. Before this analysis was performed a post hoc decision was made to investigate the relationship between Accuracy Scores and the interaction of a person's field D/I score and social intelligence. This decision was based upon the speculation that persons who are both field independent and high in social intelligence may be relatively superior in accuracy than those persons who do not possess both traits. To this end a new variable was formed, field D/I x social in-

⁵This issue is explicitly discussed in a later section.

telligence, by multiplying a subject's field D/I score with his/her social intelligence score (Ezchiel & Fox, 1959, Ch. 21)⁶. This new variable was found to significantly correlate with accuracy ($r = .274$, one-tailed $p < .003$). In addition, a post hoc decision was made to correlate the interaction of social intelligence and general intelligence with accuracy. This follows the suggestion of Taft (1955, p. 20) that it may be a combination of these traits that are most related to accuracy in person perception. The positive correlation of this new variable with accuracy ($r = .239$, one-tailed $p < .008$) supports the post hoc hypothesis that persons who are high in both general intelligence and social intelligence are superior in accuracy to those persons who are relatively low in both social intelligence and general intelligence.

Taking these results into consideration the aforementioned regression analysis was performed using Accuracy Score as the criterion variable and the following as possible predictor variables: (a) field dependence/independence; (b) social intelligence; (c) general intelligence; (d) locus of control; (e) degree of personal moral code; (f) number of hours read per day; (g) number of hours T.V. watched per day; (h) number of Shakespearan plays read; (i) number of movies viewed per month; (j) total siblings; (k) field dependence/independence x social intelligence interaction; and (l) social intelligence x general intelligence interaction. This set of predictor variables was chosen on a post hoc basis from the results of the correlation analyses i.e., those vari-

⁶This multiplication was done after the raw scores on each variable were standardized and a constant had been added to each standard score to eliminate negative values.

ables which were found to be at least marginally related to accuracy were included.

The first variable to be chosen for inclusion was the field dependence/independence x social intelligence interaction (X_1). This variable formed Equation I⁷ that significantly predicts Accuracy Score with a multiple-R of .274 and R^2 of .075 ($F(1, 98) = 7.97, p < .006$).

$$(I) \quad \text{Accuracy Score} = .274 X_1$$

On the second step, the variable, number of hours of T.V. watched per day, was chosen for inclusion. Combined with the first variable this second variable (X_2) provided Equation II which significantly predicts Accuracy Score with a multiple R of .366 and an R^2 of .134 ($F(2, 97) = 7.49, p < .001$).

$$(II) \quad \text{Accuracy Score} = .274 X_1 - .242 X_2$$

Equation II is a significantly better predictor than Equation I, i.e., an increase in R^2 of approximately .06 ($F(1, 97) = 6.56, p < .01$).

On the third step the variable, the number of siblings the person has, was chosen for inclusion. The addition of this variable (X_3) to the first two variables provided Equation III which significantly predicts Accuracy Score with a multiple R of .421 and an R^2 of .178 ($F(3, 96) = 6.91, p < .0005$).

$$(III) \quad \text{Accuracy Score} = .276 X_1 - .246 X_2 - .209 X_3$$

⁷Regression coefficients are in standardized form.

Equation III is a significantly better predictor than Equation II, i.e., a significant increase in R^2 of approximately .04 ($F(1, 96) = 5.11$, $p < .02$).

On the fourth step the variable, moral code, was chosen for inclusion. Combined with the other three variables, this variable (X_4) provided Equation IV which significantly predicts Accuracy Score with a multiple R of .439 and an R^2 of .192 ($F(4, 95) = 5.66$, $p < .0006$).

$$(IV) \quad \text{Accuracy Score} = .264 X_1 - .227 X_2 - .195 X_3 + .125 X_4$$

Equation IV is not a significantly better predictor than Equation III, i.e., a nonsignificant increase in R^2 of approximately .015 ($F(1, 95) = 1.76$, n.s.). Therefore, Equation III appears to be the best and most parsimonious predictor of an Accuracy Score. Inspection of Equation III shows that persons who are both field independent and of high social intelligence, spend little time watching T.V., and have few siblings are high in Accuracy. On the other hand, persons who are field dependent and low in social intelligence, spend relatively more time watching T.V., and have many siblings are relatively lower in Accuracy. In addition it can be noted that the standardized regression coefficients in Equation III are similar in absolute size. This indicates that the three variables (field D/I x social intelligence interaction, number of hours of T.V. watched per day, and number of siblings a person has) assume relatively equal importance in predicting Accuracy.

Inspection of the correlation results in Table 5, shows that the degree of confidence a subject displayed in his/her rating is also significantly correlated with Accuracy Score; in addition Confidence Score

appears to be relatively independent of the three subject variables in Equation III. As such, a post hoc decision was made to form a new predictor equation by adding the Confidence Score variable (X_5) to the variables in Equation III. This new Equation (V) predicts Accuracy Score with a multiple R of .466 and an R^2 of .217 ($F(4, 95) = 6.57, p < .0002$).

$$(V) \quad \text{Accuracy Score} = .269 X_1 - .251 X_2 - .201 X_3 + .198 X_5$$

This equation is a significantly better predictor than Equation III, i.e., an increase in R^2 of approximately .04 ($F(1, 95) = 4.75, p < .03$).

A more complete perspective on Equation V's efficiency in predicting accuracy is presented in Figure 4. This figure is a scatterplot of actual Accuracy Scores and predicted accuracy scores⁸. This figure, in addition to providing a visual display of the regression results ($R = .466$), allows for an inspection "how efficient" the equation (Va) is in predicting specific subjects' Accuracy. For the purposes of this analysis it was decided to refer to subjects with an Accuracy Score of 13 as a "significantly" accurate subject, i.e., a person who is approximately 2/3 correct in his/her judgment (see Appendix I for further discussion of the rationale underlying this decision). Figure 4 has been divided into four regions for further understanding of Equation Va's efficiency.

⁸A predicted accuracy score was computed for each subject using the following equation (regression constant are in nonstandardized form:

$$(Va) \quad \hat{AS} = .04267 X_1 - .41253 X_2 - .20534 X_3 + .07034 X_5 + 5.34865$$

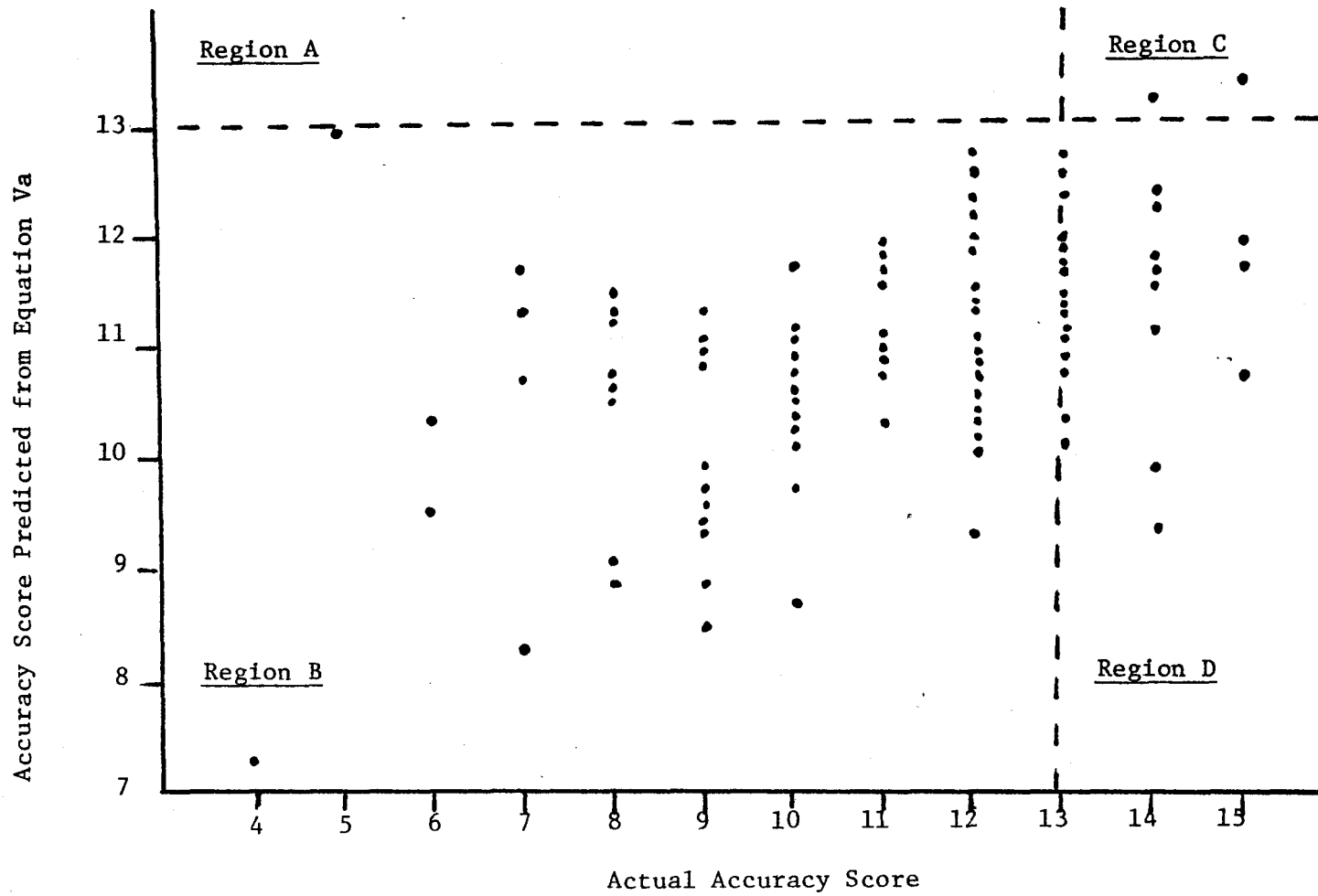


Figure 4. Scatterplot of actual accuracy vs. predicted accuracy.

- (a) Region A identifies an area of False Positives, i.e., a person is predicted as significantly accurate when in fact he/she is not. No points fall within this region (but one comes close).
- (b) Region B identifies an area of True Negatives, i.e., a person is predicted as having an Accuracy Score below the significant cut-off (AS = 13) and in fact this person's actual Accuracy Score is less than 13. Seventy-one percent of the points fall within this region.
- (c) Region C identifies an area of True Positives, i.e., a person is predicted as having an Accuracy Score equal to or above the significant cut-off and in fact the person's actual Accuracy Score is a significant one. Two percent of the points fall within this region.
- (d) Region D identifies an area of False Negatives or Type II Errors, i.e., a person is predicted as having an Accuracy Score less than the significant cut-off, when in fact his/her actual Accuracy Score is a significant one. Twenty-seven percent of the points fall within this region.

These results indicate that Equation V is a good predictor when it comes to avoiding False Positives (i.e., no one was falsely predicted as being significantly accurate), but not a good equation when it comes to False Negatives (i.e., 27 of the 29 subjects who were significantly accurate were predicted to be not significantly accurate). The implications of these findings are discussed later.

Trust Scores. Due to the lack of support for the hypothesized relationships, the exception being the low but significant correlation

between the Trust Score and Trustworthiness of Others, it was deemed most appropriate not to pursue any multiple regression analyses. This decision was made, given the consideration that while some significant post hoc results may occur, the lack of findings at the bivariate correlational analysis step did not lend support for such further aposteriori analyses.

Discriminate Analysis for Accuracy Scores

In order to further explore the relationships between the subject variables and Accuracy, a stepwise discriminant analysis was performed. The specific purpose of this analysis was to determine what subset of predictor variables would most accurately classify subjects into one of two groups. Group One represents those subjects whose Accuracy Score is below the significant cut-off value of "13" and Group Two represents those subjects whose Accuracy Score was equal to or above "13". To this end, a new variable was generated with a value of "1" assigned to those subjects whose Accuracy Score was less than "13" and a value of "2" assigned to those subjects whose Accuracy Scores was greater than or equal to 13. The discriminant analyses was then performed to find a discriminant function which would optimally predict whether a subject was a member of Group One or Group Two.

In this analysis the same subset of subject variables used in the stepwise multiple regression analysis was employed. A computer program, SPSS VI Discriminant, was programmed to perform the analyses in accordance with the Wilke's criterion (Λ) for solution⁹. In addition the

⁹ Wilke's criterion (Λ) is the overall multivariate F ratio for the test of differences between group centroids (Klecka, 1975, p. 447).

computer was informed that of the total 100 subjects, 71 actually were in Group One and the other 29 were in Group Two. Table 6 displays the differences in group means for the variables in the predictor subset. As shown in this table, nine of the variables have significantly ($p < .03$) different groups means¹⁰. These significant differences indicate that compared with Group One persons, Group Two persons (the significantly accurate group) are more field independent, have more social intelligence, have a higher general intelligence, are more likely to have a personal moral code on which they base everyday decisions, watch less T.V. per day, have read more Shakespearan plays, watch more movies per month, are more field independent and socially intelligent, and are more socially intelligent and generally intelligent. Using this subset of variables the discriminant analysis program found a discriminant function which included three predictor variables. These variables (the number of movies watched per month, the number of Shakespearan plays read, and the interaction of social intelligence and general intelligence) combine to significantly discriminate between Group One and Group Two, (Wilkes' $\Lambda = .88$, $\chi^2(3) = 12.124$, $p < .007$); and have standardized discriminant function coefficients of $-.559$, $-.384$, and $-.552$ respectively. The efficiency of this discriminant function to accurately predict whether a subject is in Group One or Group Two is shown in Table 7. This table shows that 78% of the subjects were correctly classified into whether

¹⁰In interpreting this finding, the disadvantages of multiple t-tests are important to remember. The probability that at least one of these 12 t-tests will be spuriously found to be significant is:

$$1 - (1 - .03)^{12} = .31 \text{ (Hays, 1963, p. 367).}$$

Table 6

Group One and Group Two Means for
Subject Variables in Discriminant Analysis

Variable	Gr. One \bar{X}	Gr. Two \bar{X}	<u>t</u> of Diff	p<
Field D/I	66.34	72.55	-2.57	.01
Social Intelligence	22.97	24.59	-2.66	.009
General Intelligence	15.27	16.31	-2.48	.01
Locus-of-Control	11.01	10.79	_____	n.s.
Moral Code	6.92	7.52	-2.40	.02
Hours Read	2.51	2.69	_____	n.s.
Hours TV	1.89	1.48	2.16	.03
# Plays	3.69	5.41	-3.09	.003
# Movies	1.35	1.86	-3.33	.002
# Siblings	2.94	2.31	_____	n.s.
F D/I x SI Interaction	-2.07	5.41	-3.49	.001
SI x GI Interaction	-1.94	5.61	-3.37	.001

Table 7

Efficiency of Discriminant Function

Actual Group	Number of Subjects	Predicted Group One	Predicted Group Two
Group One Non-significant	71	69 (97.2%)	2 (2.8%)
Group Two Significant	29	20 (69%)	9 (31%)

Note. Percentage of "grouped" subjects correctly classified: 78%

they were or were not significantly accurate in their ability to rate the tapes. Further inspection of Table 7 indicates that 69% of the subjects actually in Group Two were falsely classified into Group One (i.e., False Negatives), while 2.8% of those subjects actually in Group One were falsely classified into Group Two (False-Positives). The efficiency of this discriminant function, although based on a different set of subject variables than the regression equation (V) is similar to the efficiency of Equation V in that it avoids False Positives, but not False Negatives (i.e., it identifies only two subjects as being significantly accurate when in fact they were not, while it identifies 20 subjects as not being significantly accurate when in fact they were).

DISCUSSION

Judgmental Accuracy

Consistent with past findings (Hocking, Note 1; Maier, 1966; Maier and Jansen, 1967; Maier and Thurber, 1969; and Maier and Lavrakas, 1976), and as hypothesized, listeners as a group displayed a better-than-chance ability to judge veracity in the audio medium. While this above-chance accuracy is significant, the percentage of correct judgments was only 54.45%, which is considerably lower than the 77% correct Maier and Thurber observed in their listeners group; in fact, of the 100 subjects in the present study, only four performed at a level comparable of the average listener of Maier and Thurber. Another study to report the average percentage of correct judgments by "listeners" is the investigation by Hocking (Reference Note 1). His audio-only group displayed averaged judgmental accuracy of 62% correct.

The fact that the listeners in the present study were not as accurate as listeners in these other studies may well be due to the present operationalization of the veracity construct. Listeners in the present study had a relatively short exposure (approximately 10-15 seconds) to each stimulus person's lie or truth. This not only limited the length of time the stimulus-person spoke, but also limited the quantity of verbal information the stimulus person presented. On the other hand, listeners in other studies were exposed to stimulus persons for a longer time duration and presumably heard the stimulus-persons present more verbal information. While this line of reasoning is speculative,

it follows from an assumption that in general the more spoken information a listener has (up to some saturation level) the more able the listener will be to judge accurately. I do not regard this reasoning as inconsistent with the "information-overload" hypothesis advanced by Krauss et.al. (Reference Note 2) to account for judges' relative inferior accuracy when judging audio and visual information. In that instance, the visual information is thought to "distract" the untrained observer; and, as Ekman and Freisen (1969) suggest, the face is well-equipped to lie successfully.

Also consistent with past findings is the variation in judgmental accuracy shown by the present listeners. As was found in the pilot study, judgmental accuracy in the present study ranged from approximately 20% to 75% correct, with a negatively skewed observed frequency distribution of Accuracy Scores. This negative skewness indicates, as Figure 3 displays, that the observed distribution of frequency scores deviates from normality in a "more accurate" direction. This suggests that while listeners do vary in their judgmental accuracy, a clear majority of judges (70%) displayed accuracy that was equal to or better-than-chance.

This better-than-chance judgmental accuracy serves as the backdrop for the present study's primary purpose i.e., an investigation of individual's differences as they relate to listeners' accuracy in human lie detection. After careful inspection of the data and interpretation of the results of the correlation, regression, and discriminant analyses it appears that a tentative profile can be advanced regarding

"who is a good lie detector"¹¹:

A good human lie detector appears to be a person who has a field independent mode of perception, is high in social and general intelligence, has few siblings, shows an active interest in literature, films, and the manner in which he/she conducts his/her daily life, and makes his/her judgments with some degree of confidence. This description can be interpreted as follows:

1. The listener's task in human lie detection can be conceptualized as an effort to perceive the "field" (i.e., the gestalt of the speaker) as a whole made up of separate, yet interdependent, parts (i.e., the various behavioral cues of the speaker), which can be evaluated as evidence for truthfulness or for falseness. As hypothesized field independent persons are better able to accurately judge veracity. This may relate to their mode of perception which apparently leads them to experience "the parts of the field..as discrete from organized background, and [as such] deal with them separately" (Gary, 1967, p. 5190).
2. It is especially interesting to note that the interaction of high social intelligence and high general intelligence is significantly related to accuracy in human lie detection. This is in direct agreement with Taft's (1955, p. 20) conclusion that is in the combination of both forms of intelligence that enables a person to be a more "sensitive" evaluator of others. Human lie detec-

¹¹This profile is, of course, limited by the overall degree of validity of the present study, which will be explicitly discussed in the following section.

tion in this instance can be conceptualized as an "interpersonal problem solving task"; as such, being skilled in both social and general intelligence may increase the likelihood of a correct solution to the listener's "truth or lie problem".

3. Zajonc (1975) has recently advanced a theory for the relationship between birth order, family size, and intelligence. He argues that intelligence is in part a function of the amount of experience a child has had interacting with adults. The present parallel to Zajonc's reasoning is the suggestion that the more a child interacts with an adult and observes adult-behavior, the more likely he/she will become acquainted with "adult modes of judging others" (Taft, 1955, p. 8). The finding that having few siblings is positively related to judgmental accuracy in human lie detection, can be interpreted following this child-adult hypothesis. In general, children with fewer siblings are more likely to spend a larger proportion of their time interacting with adults than are children with relatively more siblings; therefore the child with few siblings may more likely be exposed to adults qua perceivers in human lie detection, and thereby become more familiar with the adult-approach to "detecting" the lies of others.
4. Taft (1955) suggested that esthetic-literary interests and abilities were very likely to be correlates of the ability to judge others accurately. While the scores on the improvised Artistic Interest Scale did not relate to accuracy in the present study other variables which may be part of the esthetic-literary

construct did show positive relationships to accuracy: (a) reading relatively more hours per day; (b) watching relatively less hours of television per day; (c) watching relatively more movies at a theater per month; and (d) having read relatively more Shakespearan plays. If these can be interpreted as indicating esthetic-literary experiences then it may be such that a person's superior accuracy in judging veracity is related to his/her being a "broader" individual, i.e., one who is more familiar with literary-esthetic interpretations of human emotion. As none of these variables show any strong correlation with social intelligence or general intelligence they appear to represent a dimension independent of "intelligence".

5. It was found that a person who has a strong personal moral code by which he/she makes everyday decisions is more likely to show superior judgmental accuracy. It can be speculated that the person with a strong personal moral code will be more aware of the "morality" of others, than will a person without a strong personal moral code. If such a person is aware of the morality of others, he/she may well be "constantly" alert to the possibility that another person is lying and as such become more capable of detecting such behavior.
6. Finally, the good lie detector has a relatively greater degree of confidence in the judgments he/she makes. While Maier and Jansen (1966) suggested that judges may not be able to articulate the reasons for their judgments, it does appear that the listeners in the present study had some degree of "self-know-

ledge" regarding their judgment. This suggests that accurate judges "know" they are accurate while less accurate judges "realize" they are uncertain.

In general, this description of a "good human lie detector" corresponds, in part, with my a priori hypotheses and also shows some correspondence with the description of a person who is, in general, a good judge of others (cf. Taft, 1955, 1956). This is not surprising as the listener qua perceiver in human lie detection is obviously engaging in a form of person perception, i.e., is the speaker to be perceived as a truth-teller or a liar?

Of additional interest are the findings regarding the efficiency of the prediction equations. As earlier mentioned, both the multiple regression equation and discriminant function were efficient in that they did not predict False Positives. If subsequent research validates the present predictive relationships and their efficiency, this lack of predicting False Positives would be important for anyone who may want to identify "good human lie detectors" with a low probability of predicting incorrectly.

Contrary to the set of a priori hypotheses, judgmental accuracy was not found to relate to extraversion, neuroticism, locus-of-control, or dogmaticism. It is also of interest to note that neither age nor sex were found to relate to judgmental accuracy.

In accordance with another a priori hypothesis, listeners as a group, showed a significant tendency to be more trusting, than distrusting, in their ratings. It was suggested that such behavior could result from a combination of the ambiguity of the judgment task (i.e., listen-

ers were led to believe that the tape stimuli could be any proportion of lies and/or truths) and from the fact that listeners had nothing to lose in making a mistake. This reasoning follows from an assumption that listeners are acculturated to presume innocence, and to consider it less equitable if an innocent person is mistakenly found guilty than if a guilty person is mistakenly found innocent. Based upon this interpretation of Accuracy and Trust results, I suggest that listeners in the present study were able to judge veracity better-than-chance, but in making judgmental mistakes they were more likely to err in the direction of judging the statement as true.

This interpretation of listener's judgments suggest that there may be two relatively independent components underlying a judgment in the present study, i.e., Accuracy and Trust ($r(98) = .111$, n.s.). Conceptually, this follows the aforementioned interpretation that listeners' judgments do not merely reflect Accuracy. This issue of judgments reflecting more than Accuracy appears reminiscent of Cronbach's (1955) interpretation of traditional accuracy-measures in person perception which is discussed in the following section.

Judgments in Human Lie Detection and Cronbach's Measurement Components

Cronbach (1955) criticized traditional measures in person perception arguing that the "accuracy score" these measures produced was actually made up of four components (elevation, differential elevation, stereotype accuracy, and differential accuracy). In the person perception context, the component most closely approximating "true accuracy" is differential accuracy

because the various response bias components (elevation, differential

elevation, and stereotype accuracy) have been eliminated.... The immediate effect of Cronbach's article was to render invalid or uninterpretable most of the previous research in this area.... In the period since 1955, fewer studies on accuracy in person perception have been reported. To be able to ask and answer the proper questions in this area requires extreme care in data collection and analysis of the data (Hastorf, Polefka, and Schneider, 1970, p. 32).

While Cronbach's criticism seems to have put a damper on accuracy-in-person-perception research, how does his criticism relate to the manner in which Accuracy was determined in the present study?

The Accuracy Score computed here is based upon the number of correct dichotomous judgments the listener made in his/her rating of the stimulus set. I suggest that these dichotomous judgments represent a listener's first stage in the decision process (see Figure 5) i.e., is the speaker telling the truth or lying? The second stage in this decision process involves the listener's determination of his/her degree of "certainty". While I am not suggesting that a listener implicitly thinks through these two stages, I do assert that in general listeners make their ratings following this "unconscious" reasoning¹². Had the Accuracy Scores been based on the specific responses on the 6-point scale, "true accuracy", confidence, and trust would have been totally confounded. The dichotomous manner in which the Accuracy Scores were computed eliminates the potential confound of confidence by collapsing across the confidence categories. In addition, it also eliminates the potential confound of "trust", with the possible exception of those instances in which a listener was totally uncertain and therefore

¹² A possible exception to this decision model would be an instance in which a listener cannot decide lie or truth, therefore reasoning that he/she is uncertain first, with the lie or truth "guess" coming second.

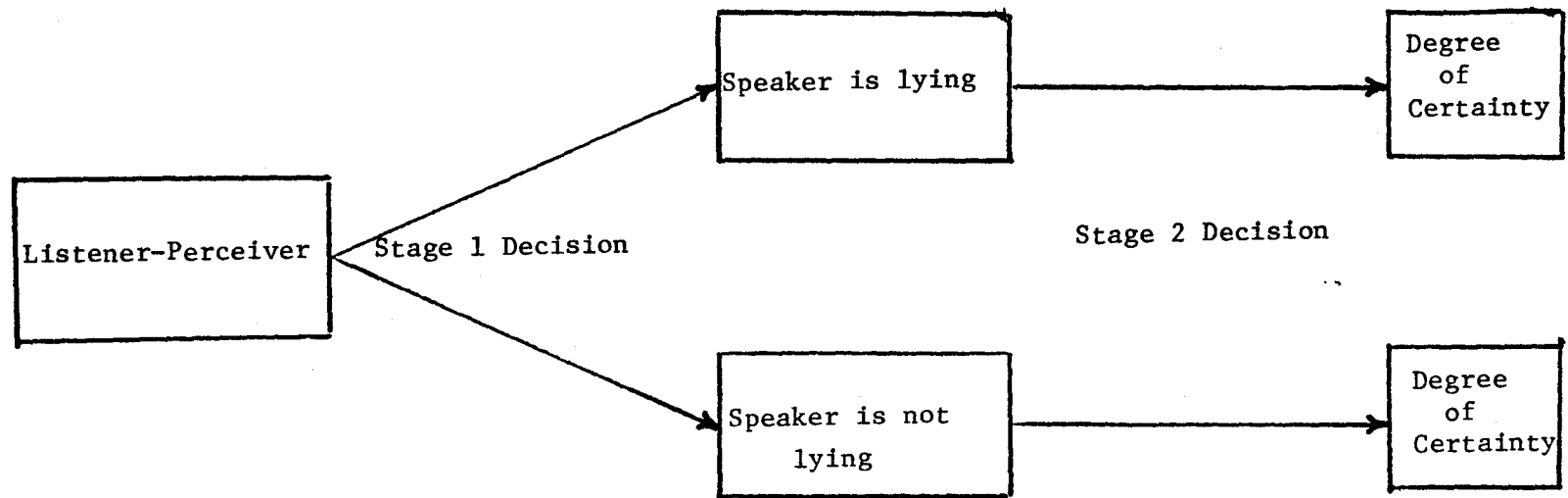


Figure 5. Two-stage judgment model for human lie detection.

"guessed" lie or truth. The results of the pilot study suggest that these instances of total uncertainty are not frequent; this was reflected by the virtual nonusage of the "totally uncertain" midpoint on the pilot study's seven-point lie/truth rating scale. All this lends support to the validity of the Accuracy Scores used in this study as relatively unconfounded measures of Accuracy. This is not meant to imply that they are totally reliable (an issue that will be later discussed), but that they are free from the confounds of confidence and trust.

On the other hand, the Trust Scores in the present study, while having some construct validity, are not totally "pure". This follows from the fact that the "true trust" component of the trust scores is confounded with accuracy and confidence. By this I mean that the actual ratings the listeners made on the 6-point scale are a mutual product of accuracy, confidence, and trust. Following the two-stage decision model proposed above, "true trust" will presumably affect a judgment when a listener is totally uncertain and therefore "guesses" lie or truth. This should be especially true when listeners face the judgment task with ambiguity as to the proportion of lies and truths in the stimulus set. Secondly, it is possible that trust affects the degree of confidence a listener shows when he/she adheres to the two-stage decision model. By this I mean that a listener who is, in general, more trusting-of-others may use the certain-response more than the suspect-response if he/she first decides "truth", or the suspect-response more than the certain-response if first deciding "lie". But at present, this remains mere speculation.

If listeners had known that the proportion of lie and truth tapes

was 50-50, or if they had explicitly taken a 50-50 judgment strategy, a "true trust" score may have been computable as a component underlying the overall ratings on the six-point scale¹³. As the listeners, here, were not given a prior expectancy, and their behavior indicates they did not follow a 50-50 strategy, a "true trust" component is not available from the raw data. Despite this problem it is interesting to note that Trust-ability of Others does correlate with the Trust Score in the present study lending some "construct" validity to the Trust Score, i.e., the "true trust" component of a listener's judgment is apparently contained within the Trust Score.

In summary, the validity of Trust Scores is uncertain due to their confounding with Accuracy and Confidence. But of critical importance for the present study is the assertion that Accuracy Scores as computed here do reflect unconfounded Accuracy.

Validity Issues

Statistical Conclusion Validity. The data analyses in the present study were performed in both an a priori and a posteriori manner. Specifically (with the exception of the one sample t-tests and bivariate correlational analyses) the step-wise multiple regression and step-wise

¹³If a 50-50 strategy had explicitly been used subjects' deviation from a Trust Score of 70 would be a "purer" indication of true behavioral trust, but not a perfect one. For a relatively unconfounded measure of true behavioral trust, as reflected in judgments, an investigator would need to know on which judgments listeners were totally uncertain. Then by observing which "side" of the lie/truth rating scale the totally uncertain subject used, the investigator could form a "true trust score" which should be more reflective of the subject's general level of behavioral trust of others. As these possible instances of total uncertainty are indeterminable, in the present study, such an "unconfounded" measure of true trust is not available.

discriminant analysis are done after-the-fact. That is, I determined the subset of variables to be entered as predictors for these multivariate techniques on the basis of the bivariate correlation results. In addition, these techniques themselves are "post hoc" in a sense that they produce optimal results, i.e., "they take what you give them, and give you back the best that's there". With this in mind, it is possible that some Type I Errors are present, given the multitude of individual tests performed. But due to the variety of statistical tests performed I cannot estimate what the experiment-wise error rate (Kirk, 1968, p. 83) of the entire study is. Despite this limitation, strong support of the suggestion that the multivariate results are not spurious comes from the very low levels of probability associated with them (the regression equation had a significance level of less than .0004 and the discriminant function had a significance level of less than .006). Given this, and the size of the sample (large enough to minimize Type II Errors), I am led to the conclusion that the results presented here have good statistical conclusion validity. None-the-less, replication is needed before this question will receive a more definitive answer.

Internal Validity. As the present study was not designed to test cause-and-effect relationships, concern for internal validity becomes one of "what validity do the interpretations of the observed relationships have?" In proposing a "tentative profile" of the good human lie detector, possible cause-and-effect interpretations have been advanced that could, in my opinion, account for the covariation between field independence, social intelligence, etc. and judgmental accuracy. As these possible causal relationships were not tested by the present study

they are no more than plausible explanations for the observed covariation. The "validity" of these plausible explanations must then be judged against the reasonableness of the logic they are based upon. While the explanations that are advanced for the observed relationships between field independence, social intelligence, and general intelligence are based on what I consider to be sound conceptual reasoning, the explanations for the observed relationships between accuracy and the literary-esthetic experience, number of siblings, and confidence are speculative. Despite this self-assessment, the answer to the question of the internal validity of the "causal-interpretations" must await experimental research.

Construct Validity. In order to put the results of the present study in a proper perspective one must consider the construct validity of the criteria and predictor variables. The predictor constructs, for the most part, were measured by tests with recognized reliability and validity. A criticism could be advanced that use of abbreviated inventories for measuring field dependence/independence and general intelligence lessens confidence in the validity of these scores. This criticism may be correct from a reliability standpoint, i.e., shortening the tests may have reduced their reliability and hence validity. If this is the case, then the trust of this criticism should be directed in the Type II Error direction, i.e., a more reliable measure may have found a

a stronger relationship between Accuracy and these variables¹⁴. In addition to these possible criticisms, an obvious criticism is the lack of reliability information regarding the Artistic Interest scale. This is enough of a problem to render totally uninterpretable the finding of no relationship between Accuracy and artistic interests. Finally, the questions on the demographic and background questionnaire have no proven reliability. While this should not be a concern for the descriptive items, e.g., age, sex, number of siblings, etc., there is uncertain reliability for the personal habits and personal attitudes questions. Again, following the reasoning advanced earlier, if these questions produced unreliable measures of the predictor constructs they were intended to tap, then it is possible that some Type II Errors have been made, i.e., it is possible that some predictor constructs and accuracy are related, but the tests of these relationships may not have been sufficiently sensitive.

While discussion was earlier presented regarding whether Accuracy Scores reflect "true accuracy", the reliability of these Accuracy Scores was not discussed. As listeners in the present study did not return to make a second set of judgments on an equivalent stimulus set, there is no information regarding reliability over time. This lack of "test-re-

¹⁴This follows from the correction for attenuation (Guildford, 1943, p. 400):

$r_{yx_t} = \frac{r_{yx}}{r_{xx}}$ where r_{yx_t} represents the true correlation of x and y, r_{xy} , represents the observed correlation between x and y, and r_{xx} represents the reliability of x.

test" reliability information is also the case for the other studies of perceiver accuracy in human lie detection, which were reviewed in the introduction section of this dissertation. In addition, none of these studies have reported the internal consistency of their accuracy scores, i.e., what is the interrelationship among the judgments made on the different stimulus persons?

Hunt and Lin (1967) had found that accurate judges were accurate across different personality-types, which suggests that a "good" judge is good in a general sense. Despite this general ability that Hunt and Lin's good judges showed, it is quite possible that different good judges were accurate on different subsets of judgments. As Hunt and Lin and the human lie detection studies reviewed earlier did not report any of these results, this question is unanswerable as it applies to these investigations.

In the present study this issue becomes one of asking whether individual judges "earned" their Accuracy Scores by being accurate on exactly the same subset of stimulus persons. If this had been the case, one would expect some tapes to be rated correctly nearly 100% of the time and others to be rated correctly nearly 0%. The observed range of percentage of judgments correct for the stimulus tapes was 32%-77%. This range, while not as extreme as 0%-100%, does not rule out the possibility that judges were consistently good or bad on the same tapes. To answer this question more precisely a between-persons homogeneity

coefficient¹⁵ was calculated. This value for the present 2,000 judgments is .181; which means that 18% of the variation in Accuracy Scores, as measured by the stimulus set in the present study, is accounted for by actual individual differences in accuracy. How should this value be interpreted?

In interpreting this coefficient it is important to remember that the set of stimulus tapes represents 20 different speakers each in one of 20 different speaker conditions. It therefore follows that these 20 stimulus persons represent 20 different lies or truths. While these different stimuli are in one sense qualitatively similar, it is evident by the "low" value (.181) that the tapes are judged with differential accuracy, e.g., listeners with an Accuracy Score of 11 apparently made their 11 correct judgments on different subsets of 11 tapes. This "low" homogeneity coefficient indicates a good deal of heterogeneity among listeners on which tapes they were accurate. This suggests that the Accuracy Score in this study is made up of as many as 20 "different kinds of accuracy"¹⁶. Thus, an Accuracy Score can be interpreted as reflect-

¹⁵The following homogeneity coefficient indicates the amount of variance in accuracy that is directly accounted for by consistent individual differences in accuracy among the 20 judgments (Guildford, 1954, p. 384):

$$\text{between-persons homogeneity} = \frac{MS_{bp} - MS_{res.}}{MS_{bp}}$$

¹⁶In fact, a factor analysis of the intercorrelation of the 20 "accuracy" scores identified nine factors accounting for 64.5% of the variance. Yet a Varimax rotation produced loadings that were not meaningfully interpretable regarding the conditions that were used to generate the 20 tapes.

ing general accuracy across the set of tapes i.e., the higher a listener's Accuracy Score the more accurate he/she was in judging a variety of lie and truth conditions.

The construct validity of the Trust Scores in the present study has been earlier discussed. At that time, it was stated that a listener's degree of behavioral trust in others (as measured by Trust Score) is confounded with accuracy and confidence. As such, Trust Scores are a low validity measure of behavioral trust as they apparently contain other sources of variation other than trust and error.

External Validity. The generalizability of the tentative description of a good human lie detector must necessarily be limited by the situational specifics of the judgmental task employed. By this, it is meant that the judgmental accuracy measured here is the ability to listen to the speech of someone who is "voluntarily" advocating an honest or dishonest position in order to "help" someone else (the experimenter), and then accurately judge whether the speaker is telling the truth or lying. In addition, listeners had "nothing to lose" by making a mistake in judgment. This raises the question of motivation level of the listeners. Would similar predictor-criterion relationships have been observed if there had been some source of extrinsic motivation, e.g., a reward for the listener with the highest accuracy? In the present study, listeners' motivation to "do the best they could" may have been totally intrinsic; listeners who wanted to help the experimenter by performing well may have been the most motivated to concentrate on the judgment task. If differential motivation is a factor in the accuracy of listeners it may as well confound the observed predictor-criterion

relationships. While this could be a potentially serious limitation of the findings, it should be remembered that all listeners were informed of the importance of this research due to the implications it might have in future legislative decisions. If listeners regarded these introductory instructions with the sincerity and importance with which they were spoken, it may be justified to assume that the listeners were extrinsically motivated to do their best. Unfortunately this issue must remain unresolved.

In addition to the motivation level and specific conditions under which listeners heard the lies and truths, a proper limitation to the generalizability of these findings is the population of humans that the listeners represent. While the observed predictor-criterion relationship might hold for all humans, the present findings must be limited to a population of relatively well educated, primarily Caucasian, urban-middle class, young adult males and females. This limitation of generalizability is necessary until this research is replicated on different age and social-economic-status level individuals.

Finally, the question of the generalizability of the "lies and truths" themselves must be discussed. In the present study a speaker was voluntarily asked to help another person (the experimenter) by providing an honest or dishonest testimony. From the speaker's perspective the testimony would, at most, cause another person to either waste a few minutes participating in a dull task or miss participating for a few minutes in an interesting task. I suggested these lies and truths have mundane realism due to the assumption that many humans engage in such "persuasion" on a regular daily basis. On the other hand,

the stimulus persons in the present study, did not confront the listeners for whom their testimony was allegedly intended. It may be assumed that Ekman and Friésen (1969) might not regard these "lies" to be totally valid deceptive communications according to their criterion dimensions of deceptive interactions: (a) there is an uncertain saliency to the deception; (b) the stakes for the speaker to be successful are somewhat low; (c) there is no listener to confront the speaker, and therefore no balance of roles; and (d) there is no struggle between deceiver and potentially deceived about the maintenance of the deception. While the speakers' "lies" in the present study are deceptive communications in a broad sense, they do not closely conform to Ekman and Freisen's criteria. Whether this is a valid limitation to the generalizability of these findings is at present uncertain.

A final issue regards the ethics of the methodology used to generate the stimulus lies and truths. Conforming with the four criterion dimensions, Ekman and Friesen (1974) and Hocking (Reference Note 1) generated a set of stimulus tapes on the assumption that it was correct and ethical to inform their subject-speakers that to be a success in their chosen field, they themselves might have to become masters of the deceptive communication. In generating these stimulus tapes I did not want to "deceive" subject-speakers to such an extent. In my mind, I went as far in deception as I personally felt comfortable; I say this not as a criticism of the ethics of Ekman, Friésen, and Hocking, but merely to point out the apparent difference between myself and these

other researchers¹⁸.

Possible Future Directions for Research in Perceiver-Differences in Human Lie Detection

The present exploratory study is not intended to provide a definitive statement regarding "who is a good lie detector" i.e., what psychological dimensions validly predict human accuracy in human lie detection. Instead, this study was conceived as a necessary first step in the systematic investigation of perceiver differences. As has been reviewed, most past research has focused on stimulus properties and/or stimulus-perceiver interactions. As there appeared to have been a dearth of findings regarding perceiver differences the present study was a fruitful one for further hypotheses generation. Yet only through follow-up research can the tentative profile of a "good human lie detector" be validated.

The present findings also provide a possible comment on the dynamics of human lie detection. The listener's task in the present study was to listen to an unseen stranger and make a decision of "lie" or "truth". In the real world the majority of our verbal exchanges probably is done in the presence of another with whom we are acquainted and whom we also can "see". (An exception to this availability of visual information would be telephone conversations.) If the listeners in the present study did make judgments in accordance with the two-stage decision-model presented earlier, the judgments were free from the "bias"

¹⁸Personally, while I am uncomfortable about the deception I engaged in, I justify it by knowledge of the debriefing the subject-speakers received. In addition, I do not agree with the assumption that nurses and law enforcement officers need to be deceptive to be "successful".

from prior-knowledge of the speaker and of visual information; listeners in the present study did not approach each judgment with some predisposition towards the speaker. In the real world of the face-to-face verbal exchanges humans undoubtedly do have some such predispositions. The results regarding Differential Trust suggest that listeners were more suspicious of the stimulus persons with more "deviant" voices. If such a bias does exist outside the laboratory (and I have little common sense reason to doubt its existence) then it attenuates the decision-making model in Figure 5, by serving as a source of unaccounted for variance. Thus, while a listener might first make the lie or truth decision, any bias that he/she brings into the interaction probably affects the probability of a specific lie or truth decision; and undoubtedly affects the confidence with which the decision is ultimately made. Therefore, until more "real" world investigations of human lie detection are performed the validity of the present perceiver decision-model remains uncertain.

Possible Applied Implications of Research in Perceiver Differences in Human Lie Detection.

I could provide many specific instances in which government and industry, not to mention private citizens, might possibly apply these findings if after extensive further research they "proved" valid. But I prefer to limit my speculations to a conservative comment and a possibly controversial one.

As previously mentioned, a recent U.S. Congressional Committee has recommended legislation for the total elimination of U.S. governmental use of mechanized lie detectors. If such a course of legislative action

is enacted, governmental decisions regarding veracity would stand in the province of human decision makers, as is presently the case in trial by judge or jury. Relevant to this, the finding of this dissertation can be interpreted as lending some validity to the jury qua group approach to accuracy in human lie detection. If individuals' (qua individuals) ability to accurately judge veracity is heterogeneous across different speakers, then a group of judges should be more likely to have some members who are accurate in judging veracity. While this does not ensure that the accurate judges will convince their fellow group members of the accuracy of their perception, it does suggest that at least the possibility does exist. On the other hand, if the judges had shown homogeneity across different speakers, it would imply that even groups of judges would be unlikely to have accurate individual group members in perceiving certain types of lies or truths. Obviously the findings presented here do not provide an answer to this issue; they are merely suggestive for further research.

The more radical speculation that I advanced parallels R. Heinlein's concept of a "Fair Witness" in Stranger in a Strange Land. Heinlein (1961) wrote:

Jubal looked pained. 'You know how Fair Witnesses behave.'

'Well, no,...I don't, I've never met one,' [said Jill].

'So? Anne!'

Anne was on the springboard; she turned her head, Jubal called out, 'That house on the hilltop--can you see what color they've painted it?'

Anne looked, then answered, 'it's white on this side.'

Jubal went on to Jill, 'You see? It doesn't occur to Anne to infer

that the other side is white, too. All the kings horses couldn't force her to commit herself...unless she went there and looked---and even then she wouldn't assume that it stayed white after she left.'

'Anne is a Fair Witness?'

'Graduate, unlimited license, admitted to testify before the High Court...' (p. 64).

In the America of Heinlein's story, individuals have been licensed because of their ability (probability through training) to visually perceive a situation and describe what they have perceived free from the bias of inference. Here then is a society that "professionalizes" its witnesses and thereby places greater credulance on their perception than on the perception of the average "untrained" witness. Might we reach a time in the future in which "professional" human lie detectors serve as jurors instead of "average" citizens? If we do, for good or for bad, it may in part result from the findings of intensive research regarding the validation of who and what makes a good human lie detector.

REFERENCE NOTES

1. Hocking, J.E., Bauchner, J.E., Kaminski, E.P. & Miller, G.R. Defecting deceptive communication from verbal, visual, and paralinguistic cues. Paper presented at the International Communication Association Convention, Portland, Oregon, April, 1976.
2. Krauss, R.M., Geller, V., & Olson, G. Modalities and cues in the detection of deception. Paper presented at the American Psychological Association Convention, Washington, D.C., September, 1976.
3. Maier, R.A. Personal Communication, January, 1976.

REFERENCES

- Anderson, N.D. Likableness ratings of 555 personality-trait words. Journal of Personality and Social Psychology, 1968, 9, 272-279.
- Allport, G.W. Personality: a psychological interpretation. New York: Holt, 1937.
- Baskett, G.D. & Freedle, R.O. Aspects of language pragmatics and the social perception of lying. Journal of Psycholinguistic Research, 1974, 3, 117-131.
- Bem, D.J., Wallach, M.A., & Kogan, N. Group decision-making under risk of aversive consequences. Journal of Personality and Social Psychology, 1965, 1, 453-460.
- Bronfenbrenner, V., Harding, J., & Gallwey, M. The measurement of skill in social perception. In D. McClelland (Ed.), Talent and Society. New York: Van Nostrand, 1958.
- Byrne, D. The repression-sensitization scale: rationale, reliability, and validity. Journal of Personality, 1961, 29, 334-349.
- Byrne, D. Repression-sensitization as a dimension of personality. In B.A. Maher (Ed.) Progress in Experimental Personality Research: Volume 1. New York: Academic Press, 1964.
- Cattell, R.B. & Cattell, A.K.S. Culture fair intelligence test: scale 2. Champaign, Ill.: Institute for Personality and Ability Testing, 1960.
- Christi, R. Machiavelianism scale in unpublished manuscript. Department of Social Psychology, Columbia University, 1968.
- Cline, V.B. & Richards, J.M. Accuracy of interpersonal perception--a general trait? Journal of Abnormal and Social Psychology, 1960, 60, 1-7.
- Cronbach, L. Processes affecting scores in "understanding others" and "assumed similarity". Psychological Bulletin, 1955, 52, 177-193.
- Deutsch, M. Trust, trustworthiness, and the f-scale. Journal of Abnormal and Social Psychology, 1960, 61, 138-140.
- Ekman, P. & Friesen, W.V. Nonverbal leakage and clues to deception. Psychiatry, 1969, 63, 88-106.

- Ekman, P. & Friesen, W.V. Detecting deception from the body and the face. Journal of Personality and Social Psychology, 1974, 29, 288-298.
- Ekman, P. & Friesen, W.V. Body movement and voice pitch in deceptive interaction. Semiotica, 1976, in press.
- Eysenck, H.J. The maudsley personality inventory. London, University of London Press, 1962.
- Ezekiel, M. & Fox, K.A. Methods of correlation and regression analysis: linear and curvilinear. New York: Wiley & Sons, 1959.
- Gary, M.L. Field dependence and susceptibility to social influence (abstract). Dissertation Abstracts International, 1968, 28(12-B), 5190-5191.
- Giffin, K. The contribution of studies of source credibility to a theory of interpersonal trust in the communication process. Psychological Bulletin, 1967, 68, 104-120.
- Hastorf, A.H., Schneider, D.J., & Polefka, J. Person perception. Reading, Mass.: Addison-Wesley, 1970.
- Hays, W.L. Statistics. New York: Holt, Rinehart, and Winston, 1963.
- Heinlein, R.A. Stranger in a strange land. New York: Berkeley Publishing Corporation, 1961.
- Hunt, R.G. & Lin, T.K. Accuracy of judgments of personal attributes from speech. Journal of Personality and Social Psychology, 1967, 6, 450-453.
- Klecka, W.R. Discriminant Analysis. In N.H. Nie et al (Ed.), SPSS (2nd Ed.), New York: McGraw-Hill, 1975.
- Knapp, M., Hart, R., & Dennis, H. An exploration of deception as a communication construct. Human Communication Research, 1974, 1, 15-29.
- Lavrakas, P.J., Buri, J., & Mayzner, M. A perspective on the recognition of other-race faces. Perception and Psychophysics, 1976, in press.
- Leventhal, H. Cognitive processes and interpersonal prediction. Journal of Abnormal and Social Psychology, 1957, 55, 176-180.
- Lykken, D. The right way to use a lie detector. Psychology Today, 1975, August, 56-60.

- Maier, N.R.F. Sensitivity to attempts at deception in an interview situation. Personnel Psychology, 1966, 19, 55-66.
- Maier, N.R.F., & Janzen, J. The reliability of reasons used in making judgments of honesty and dishonesty. Perceptual and Motor Skills, 1967, 25, 141-151.
- Maier, N.R.F. & Thurber, J. Accuracy of judgments of deception when an interview is watched, heard, or read. Personnel Psychology, 1968, 21, 23-30.
- Maier, R.A. & Lavrakas, P.J. Lying behavior and evaluation of lies. Perceptual and Motor Skills, 1976, 42, 575-581.
- McDonald, A.P. Internal-external locus of control. In J.P. Robinson & P.R. Shaver (Eds.), Measures of Social Psychological Attitudes. Ann Arbor: Institute for Social Research, 1973.
- MMPI testbook. Neward, N.J.: Roche Psychiatric Service Institute, 1970.
- Nightengale, H.B. Field dependence and person perception (abstract). Dissertation Abstracts International, 1972, 32(7-B), 4223-4224.
- O'Sullivan, M. & Guilford, J.P. Six factor tests of social intelligence. Beverly Hills, Calif.: Sheridan Psychological Services, 1966.
- Rokeach, M. The open and closed mind. New York: Basic Books, 1960.
- Rotter, J.B. Generalized expectancies for internal vs. external control of reinforcement. Psychology Monographs, 1966, 80(1 Whole No. 609).
- Rotter, J.B. Generalized expectancies for interpersonal trust. American Psychologist, 1971, 26, 443-452.
- Sawatzky, D.D. & Zingle, H.W. Accurate interpersonal perception and open-mindedness. Perceptual and Motor Skills, 1969, 29, 395-400.
- Schultz, R.H.K. A shortened version of the Rokeach dogmatism scale. Journal of Psychology Studies, 1962, 13, 93-97.
- Shaver, P. Authoritarianism, dogmatism, and related measures. In J.P. Robinson & P. Shaver (Eds.), Measures of Social Psychological Attitudes, Ann Arbor: Institute for Social Research, 1973.
- Snedecor, G.W. & Cochran, W.G. Statistical Methods, Ames, Iowa: Iowa State University Press, 1967.
- Strong vocational interest blank test manual. Stanford, California: Stanford University Press, 1968.

- Taft, R. The ability to judge people. Psychological Bulletin, 1955, 52, 1-23.
- Taft, R. Some characteristics of good judges of others. British Journal of Psychology, 1956, 47, 19-29.
- Thurstone, L.L. & Jeffreys, T.E. Concealed figures: form a. Chicago: University of Chicago Press, 1956.
- U.S. House of Representatives Committee on Government Operations House Report No. 94-795. The use of polygraphs and similar devices by Federal agencies. Washington, D.C.: U.S. Government Printing Office, 1976.
- Wilhelm, P.L. & Burns, F.D. Burns and Wilhelm lie detector operation manual. Michigan City, Indiana: Burns and Wilhelm Associates, 1954.
- Wrightsmann, L.S. Measurement of philosophies of human nature. Psychological Reports, 1964, 14, 743-751.
- Wrightsmann, L.S. Personality and attitudinal correlates of trusting and trustworthy behavior in a two-person game. Journal of Personality and Social Psychology, 1966, 4, 328-332.
- Wrightsmann, L.S. Assumptions about human nature. Monterey, California: Brooks/Cole, 1974.
- Vingoe, F.J. & Antonof, S.R. Personality characteristics of good judges of others. Journal of Counseling Psychology, 1968, 15, 91-93.
- Zajonc, R. Birth order and intelligence. Psychology Today, 1975, 8(8), 37-43.

APPENDIX A

APPENDIX A

Rating Sheets for "Screening" of Lie/Truth Stimuli

Please rate this task on each of the following scales:

- | | | |
|-----------------------|-------------------------|------------------|
| 1. very dull | 1-2-3-4-5-6-7-8-9-10-11 | very interesting |
| 2. very easy | 1-2-3-4-5-6-7-8-9-10-11 | very difficult |
| 3. very unpleasant | 1-2-3-4-5-6-7-8-9-10-11 | very pleasant |
| 4. not very demanding | 1-2-3-4-5-6-7-8-9-10-11 | very demanding |

Please rate how you felt about making the tape recording:

- | | | |
|------------------------------|-------------------------|-------------------------------|
| 1. very uncomfortable | 1-2-3-4-5-6-7-8-9-10-11 | very comfortable |
| 2. very honest | 1-2-3-4-5-6-7-8-9-10-11 | very dishonest |
| 3. very ethical | 1-2-3-4-5-6-7-8-9-10-11 | very unethical |
| 4. very tempted to
refuse | 1-2-3-4-5-6-7-8-9-10-11 | not very tempted
to refuse |

APPENDIX B

APPENDIX B

Cover Story for Subjects Making Honest/Dishonest Testimonies

"Encouraging" Cover Story

We would like to have high school students participate in the task you have just completed. Unfortunately, we cannot offer credit to them to encourage their participation. Would you please help us by making a tape recording that would encourage them to participate.

I found this experiment very interesting and challenging. All in all, it seemed well worth my time to participate and I would expect you to enjoy it too.

"Discouraging" Cover Story

We would like to have high school students participate in the task you have just completed. Unfortunately, more have signed up than we need. Would you please help us by making a tape recording that would discourage them from participating.

I found this experiment very dull and unchallenging. All in all, it seemed like a waste of time to participate and I would expect you to dislike it too.

APPENDIX C

APPENDIX C

Instructions for Improvised Tape Judgments

I'm going to play some tape recordings for you of people either lying or telling the truth. I know which tapes are truths and which are lies, and I am interested if you can accurately determine which persons are being honest and which dishonest. I will play one tape recording at a time, and after each tape you are to judge the honesty or dishonesty of the speaker.

So that you can better understand the content of the tape recordings, let me explain how these tapes were made. A person came and either participated in either an interesting or dull experiment. After finishing the experiment the person was asked to make a tape recording saying that the experiment was interesting and challenging and that they would recommend it to others, or that it was basically dull, unchallenging and that they would not recommend it to others. Those persons who participated in the interesting experiment and then said it was interesting are telling the truth, while those who participated in the interesting experiment and said that it was dull are lying. In the same way those persons who participated in the dull experiment and say that it was dull are telling the truth while those who participated in the dull experiment and said it was interesting are lying.

Your task is to tell which persons are lying and which are telling the truth. I have made many tape recordings of persons lying and telling

the truth. You will rate only a few of the tapes I have made. Please rate each tape without thinking of your ratings on the other tapes. The set of tapes you will be listening to have been randomly chosen so that it is possible that all are lies or all are truths or any combinations of lies or truths. Remember, think only of the tape itself when you make your rating. Thank you.

APPENDIX D

APPENDIX D

Instructions for Read-Tape Judgments

I'm going to play some tape recordings to you of persons either telling the truth or lying. It will be your task to try and determine whether the speaker in the tape is being honest or dishonest. These tapes were made in a manner similar to the other tapes you have heard; that is, a person participated in either a dull or interesting experiment and then made either an honest or dishonest tape recording about the experiment. An important difference in this set of tape recordings is that all of these persons were reading a script. For some persons the script was an honest one, while for others it was a dishonest one. The set of recordings that you will be listening to have been randomly chosen from a larger set of tapes: therefore these tapes could be all honest, all dishonest, or any combination of honesty or dishonesty. Remember to rate each tape independently without thinking of the other tapes. Also please remember that in this set of tapes speakers are reading a script; therefore the words they speak will be identical. Thank you.

APPENDIX E

APPENDIX E

Artistic Interests Scale

Please indicate whether you would Like (L), be Indifferent to (I), or Dislike (D) working in the following occupations:

L I D Actor	L I D Interior decorator
L I D Advertising executive	L I D Minister
L I D Architect	L I D Musician
L I D Artist	L I D Music teacher
L I D Banker	L I D Photographer
L I D Carpenter	L I D Police officer
L I D CPA	L I D Sales personnel
L I D Electrician	L I D Sculptor
L I D Engineer	L I D Veterinarian
L I D Farmer	

APPENDIX F

APPENDIX F

Demographic and Background Questionnaire

1. Sex: Female Male :
2. Age: _____
3. Racial-ethnic group: black latino oriental white other
4. # of younger brothers you have _____ # of older brothers you have _____
of younger sisters you have _____ # of older sisters you have _____
5. How religious do you consider yourself?
very religious 1--2--3--4--5--6--7--8--9 not at all religious
6. Religious affiliation: Catholic Jewish Protestant Other None
7. Do you have a personal moral code that helps you make decisions in your everyday life?
Not at all 1--2--3--4--5--6--7--8--9 Very much so
8. Marital Status: single married other
9. Average amount of reading you do daily: _____ hours
10. Average amount of television you watch daily: _____ hours
11. Average # of movies you see at a theater: _____ monthly
12. # of plays of William Shakespeare you have read: _____
13. # of years of schooling you have completed: _____ years

APPENDIX G

APPENDIX G

ANOVA for Accuracy Scores

An analysis of variance was performed to investigate the existence of any pattern in the accuracy of the ratings of the 16 non-accented tape recordings. This analysis was a 2 (read vs. impromptu) by 2 (lie vs. truth) by 2 (dull vs. interesting) by 2 (male speaker vs. female speaker) repeated measures ANOVA. The purpose of this analysis was to investigate the presence of any main effects on these four factors. The result of this ANOVA was used to make a decision whether or not to form a single Accuracy Score based on all of the ratings of the tape stimuli. These main effects are shown as follows:

<u>Source</u>	<u>SS</u>	<u>df</u>	<u>MS</u>	<u>F</u>	<u>P<</u>
read vs. impromptu	6.313	1	6.313	3.255	n.s.
lie vs. truth	76.903	1	76.903	27.359	.001
dull vs. interesting	20.099	1	20.099	9.469	.01
male vs. female	4.488	1	4.488	2.489	n.s.

These main effects represent tests between the following means:

- read (3.599) vs. impromptu (3.723)
- lie (3.445) vs. truth (3.877)
- dull (3.550) vs. interesting (3.772)
- male (3.609) vs. female (3.714)

Inspection of these results shows that subjects were significantly more confident-accurate in their ratings of truth tapes than lie tapes and interesting tapes than dull tapes. The significant difference be-

tween the lie tapes and truth tapes could well be due to subject's apparent tendency to make judgments in a "trusting" direction, i.e., of all the ratings that were made for these 16 tapes 56.62% were on the "honest" side of the scale. Therefore, since subjects were generally more likely to believe a specific tape, they were also more likely to be accurate on the honest tapes than they were in their ratings of the dishonest tapes. This confound, then, probably explains the significant main effect associated with lie tapes versus truth tapes. On the other hand, there is no ready explanation for the significant difference in accuracy between the dull and interesting tapes. Despite these "significant" results a decision was made to compute an Accuracy Score based on the ratings of all 20 tapes. This decision was made, in part, for the sake of parsimony and in part because these "significant" differences in accuracy were not large in absolute value.

APPENDIX H

APPENDIX H

Reliability of the Accent/Nonaccent Distinction

In order to support the assumption that the four tapes made by the persons with "accents" were in fact distinguishable from the other 16 "non-accented" tapes, an independent group of 30 students in a Loyola undergraduate class were asked to listen to the 20 tapes and identify the "type" of voice each speaker had. This task required these subjects to indicate whether they regarded the voice as "Midwestern White" or a "non-Midwestern White". The three "accent" tapes that had been made by Blacks were identified as being "non-Midwestern White" by 100% of the judges. The fourth tape which had been made by a male with a strong "New York Jewish" accent was identified as being "non-Midwestern White" by 67% of the judges. Each of the 16 non-accented tapes were judged to be "Midwestern White" by a majority of the raters. These results support the assumption that the four "accented" tapes are reliably distinguishable from the 16 "non-accented" tapes.

APPENDIX I

APPENDIX I

Criteria for Cut-off Point

In the pilot study it had been observed that approximately 30% of the most accurate listeners' accuracy scores fell into a region that on-the-basis-of chance would have expected to contain only the 10% most accurate. This pattern was again observed in the present data i.e., while only 13% of the Accuracy Scores would be expected by chance to equal or exceed "13", 29% of the actual Accuracy Scores fall into this region. Following from this it was decided to refer to this group who were approximately 2/3's correct or better as "significantly" accurate.

APPROVAL SHEET

The dissertation submitted by Paul John Lavrakas has been read and approved by the following committee:

Dr. Richard A. Maier, Director
Associate Professor, Psychology, Loyola

Dr. Thomas P. Petzel
Associate Professor, Psychology, Loyola

Dr. Emil J. Posavac
Associate Professor, Psychology, Loyola

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dec 17, 1976
Date

Richard Maier
Director's Signature