DETECTING DECEPTIVE COMMUNICATION FROM VERBAL, VISUAL AND PARALINGUISTIC CUES: AN EXPLORATORY EXPERIMENT

Dissertation for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
JOHN EDWARD HOCKING

1976

This is to certify that the

thesis entitled

DETECTING DECEPTIVE COMMUNICATION FROM VERBAL,
VISUAL AND PARALINGUISTIC CUES:
AN EXPLORATORY EXPERIMENT

presented by

John Edward Hocking

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Communication

Gerald R. Miller
Major professor

Date April 23, 1976

ABSTRACT

DETECTING DECEPTIVE COMMUNICATION FROM VERBAL, VISUAL AND PARALINGUISTIC CUES: AN EXPLORATORY EXPERIMENT

By

John Edward Hocking

This experiment explored how differences in the verbal and nonverbal behaviors available to persons influenced their ability to detect deception. Earlier research suggested that observers would be more accurate in judging veracity when the judgments were based on observations of persons' bodies than when based on observations of persons' heads (Ekman & Friesen, 1974). These persons were making true or false statements about their personal emotional feelings as they watched stressful films. The purpose of the present experiment was to extend and refine this research by: (1) examining observer accuracy in a situation in which persons made true or false statements about a factual event, in addition to replicating the situation in which they concealed their emotional responses; and (2) to present observers with additional categories of information, both in isolation and in combinations, and to compare any resulting accuracy differences.

Samples of both lying and "truthing" behavior were videotaped under conditions of high saliency for 16 subjects.

These tapes included eight subjects who were making true or false statements about the factual content of an event they had seen and eight subjects were making true or false statements about their emotional feelings as they watched stressful slides. These tapes were shown to 730 male and female undergraduate college student observers under 14 experimental conditions. Observers saw either a close-up shot of the subject's head, a shot of the head and body with the head blocked from view, or a shot of the head and body. They viewed the tapes in either a color or a black-and-white television format. They viewed the tapes with the audio component of the 16 subjects' behavior either present or absent. These variables created a three (head-only/body-only/head-and-body) by two (color/black-and-white) by two (audio-and-visual/visual-only) factorial design. Audio-only and transcript-only conditions brought the total number of cells in the design to 14. Observers dichotomously judged the veracity of each of the 16 subjects and also indicated their degree of confidence in the accuracy of each judgment.

For the segments in which subjects made true or false statements about a factual event, results indicated greater accuracy for head-only observers than for either the body-only or the head-and-body observers, who did not differ from each other. Audio-and-visual observers were more accurate than were observers who based their judgments on visual information only. Transcript-only and audio-only

accuracy scores did not differ although both were greater than the visual only accuracy scores.

For the segments in which subjects made true or false statements about their emotional feelings, body-only observers were more accurate than head-only observers with those in the head-and-body condition falling in between. There was no accuracy difference between the visual-and-audio observers and the visual-only observers.

For the head-only conditions of both segments there was a trend in favor of color resulting in higher accuracy scores than black-and-white. Also for both segments, observers who received audio and visual information were more confident of their judgments of veracity than those who received only visual information. Head-only observers were more confident than head-and-body observers, while body-only observers were the least confident. The transcript and audio-only observers were more confident than were the visual-only observers.

The Ekman and Friesen (1974) finding was thus replicated when the deceptive communication involved statements about emotional feelings but not when it involved statements about a factual event. Implications and qualifications are discussed. Future research is suggested.

Ekman, P. & Friesen, W. Detecting deception from the body and face. Journal of Personality and Social Psychology, 1974, 29, 288-298.

DETECTING DECEPTIVE COMMUNICATION FROM VERBAL, VISUAL AND PARALINGUISTIC CUES: AN EXPLORATORY EXPERIMENT

By

John Edward Hocking

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Communication

1976

Accepted by the faculty of the Department of Communication, College of Communication Arts, Michigan State University, in partial fulfillment of the requirements for the Doctor of Philosophy degree.

Gerald R. Miller
Director of Dissertation

Guidance Committee: Gerald R. Miller, Chairman

E B Bettinghaus

Mike Dowdle

Edward Fink

David Ralph

ACKNOWLEDGMENTS

Dr. Gerald Miller provided firm intellectual, moral, and financial support for this project. It has been a privilege to observe and work with this outstanding scholar during my three years at Michigan State. Thanks GR!

Dr. Edward L. Fink went way beyond the call of duty in contributing to whatever excellence may be perceived in this paper. I am grateful for his selfless efforts. Drs. Erwin Bettinghaus, David Ralph, and Michael Dowdle rounded out my guidance and dissertation committees. They have been very helpful on this project, as well as throughout my graduate career at Michigan State.

Joyce Bauchner was a trusted research colleague and friend throughout this project. Joyce, Edmond Kaminski, and Mark Steinberg made up the research team which ran this study. It could not have been completed without the loyal efforts of these fine people.

My friend and mentor, Dr. Cal Hylton, had little to do with this dissertation. He is, however, very much responsible for its existence. His influence on my career, and my life, is pervasive.

Finally, my parents, Julia and Ernest Hocking, deserve the biggest thanks of all. The love and devotion of these wonderful people is the most valuable thing I have. It would be impossible to overstate my gratitude.

I have been very, very lucky to have been associated with each of these people. To them, to the many instructors and students who assisted and participated in this research, and to my friends at Michigan State and Georgia, my sincere appreciation and thanks.

TABLE OF CONTENTS

PAGE

LIST OF TABLES v

LIST OF FIGURES vii

CHAPTER

I. INTRODUCTION 1

The Problem 3
Some Definitions 8
The Deception Literature 11
Summary and Research Design 29

II. METHOD 39

Overview 39
Creating the Stimulus Tapes 40
Collecting the Data 59

III. RESULTS 67

Tape 69
Accuracy 75
Confidence 86

IV. DISCUSSION 97

Accuracy 97
Confidence 109
Problems with the Study 112
Future Research 118
Summary 123

APPENDIX A. Subject Participation Invitation Letter 128

APPENDIX B. Interview Format Review Sheet 129

APPENDIX C. Subjects Post-Interview Questionnaire . 130

TABLE OF CONTENTS (continued)

		PAGE
APPENDIX D.	Guidelines for Subjects' Debriefing . .	131
APPENDIX E.	Observer Questionnaire	132
APPENDIX F.	Observer's Results Letter	140
APPENDIX G.	Analysis of Variance for Factual and Emotional Accuracy Using the Transformed Scores	142
APPENDIX H.	Correlations Between Confidence and Accuracy for Both the Factual and Emotional Segments	144
REFERENCES		145

LIST OF TABLES

TABLE		PAGE
1.	Method Used to Create Lying and Truthful Stimuli Materials in All Studies Reviewed	17
2.	Sample Size in Each Cell of the Design . . .	60
3.	Subject Sex, Actual Veracity, Number of Judgments of "True" and "False," and Mean Accuracy and Confidence Scores for Both Factual and Emotional Segments on Each Tape, and Summed	70
4.	Cell Mean Accuracy Scores, Variances, and Marginal Means for the Factual Segments .	71
5.	ANOVA for Factual Accuracy	72
6.	Cell Mean Accuracy Scores, Variances, and Marginal Means for the Emotional Segments	73
7.	ANOVA for Emotional Accuracy	74
8.	Cell Means, Sample Sizes, and Marginal Means for the Two-Way Interaction Between Shot and Tape on Factual Accuracy	76
9.	Cell Means, Sample Sizes, and Marginal Means for the Three-Way Interaction of Visaudio, Shot and Tape on Factual Accuracy	77
10.	Cell Means, Sample Sizes, and Marginal Means for the Two-Way Interaction Between Shot and Tape on Emotional Accuracy . . .	80
11.	Cell Means, Sample Sizes, and Marginal Means for the Two-Way Interaction Between Visaudio and Tape on Factual Accuracy . .	82
12.	Mean Confidence Scores, Variances, and Marginal Means for the Factual Segments .	87

LIST OF TABLES (continued)

TABLE		PAGE
13.	ANOVA for Factual Confidence	88
14.	Cell Mean Confidence Scores, Variances, and Marginal Means for the Emotional Segments .	89
15.	ANOVA for Emotional Confidence	90
16.	Comparison of the Accuracy Means and Sample Sizes for Honest and Deceptive Samples of the Ekman and Friesen (1974) Experiment with Those From the Present Experiment . . .	126
17.	ANOVA for Factual Accuracy	142
18.	ANOVA for Emotional Accuracy	143
19.	Correlations Between Accuracy and Confidence and Probability Levels in Each Cell of the Design for the Factual Segments	144
20.	Correlations Between Accuracy and Confidence, and Probability Levels in Each Cell of the Design for the Emotional Segments	145

LIST OF FIGURES

FIGURE		PAGE
1.	The Experimental Design	33
2.	Room Arrangement During the Stimulus Taping	52
3.	Two-Way Interactions Between Shot and Tape on Factual Accuracy	76
4.	Three-Way Interaction Between Visaudio, Shot and Tape on Factual Accuracy	78
5.	Two-Way Interaction Between Shot and Tape on Emotional Accuracy	80
6.	Two-Way Interaction Between Visaudio and Tape on Factual Accuracy	82
7.	Three-Way Interaction Between Visaudio, Shot and Tape on Factual Accuracy	83

Chapter I

INTRODUCTION

Lying during social interaction is relatively common in our culture. The lies may vary in seriousness from the most innocuous compliment ("white lies") to falsehoods about unambiguous matters of fact which have important consequences for both liar and victim. Write Wolk and Henley (1970), "EVERYONE lies. And the person who denies that he lies is the most egregious liar of all." (p. 1).

Given the ubiquitous nature of deceptive communication and the important role it can play in the development, maintenance, and termination of human relationships, it is not surprising that people have long been concerned, even fascinated, with detecting lying. Prescriptive advice about how to detect lying has existed from as early as 900 B.C. when a papyrus Vedas described the behavior of liars:

> He does not answer questions, or they are evasive answers; he speaks nonsense, rubs the great toe along the ground, and shivers; his face is discolored; he rubs the roots of his hair with his fingers; . . . (Trovillo, 1939, p. 849)

From the time of Christ, through the middle ages, deception was detected with the method of the ordeal in which the accused liar would have to survive various forms of cruel

torture to prove his innocence. Even today, with suspicions about "shifty-eyed" criminals, there remains a considerable folklore about the behavior of liars.

Considering the pervasiveness of deceptive communication, the historical interest, and the potential importance of detecting deception, we might expect an extensive body of extant knowledge about this phenomenon. On the contrary, little is known about lying behaviors; what they are; how (and if) they can be detected. Knapp, Hart, and Dennis (1974) recently noted that, "At this point in time, almost any systematic study of deception must be labeled exploratory" (p. 16).

Research using mechanical devices to identify physiological differences between liars and "truthers" began as early as 1897 (Lee, 1953). This research, motivated primarily by law enforcement needs, has continued to the present day and provides ample documentation that the physiological responses of individuals who are lying usually differ from those who are telling the truth (Cutrow, Parks, Lucas and Thomas, 1972). However, mechanical detection of lying has many obvious limitations and is applicable only under highly controlled conditions. Social science has made a belated entry into studying the detection of lying and has focused on observable behaviors which relate to deception. The research reported in this paper attempts to add to this small but growing body of literature on the detection of deceptive communication from observable behaviors. Specifically, the experiment

reported here was designed to answer the question: How do differences in the verbal and nonverbal behaviors available to persons influence their ability to detect deception?

The remainder of this chapter describes the problem in more detail, explicates some key terms, and critically reviews previous deception literature. The chapter concludes with an overview of the experimental design and a discussion of why and how detection accuracy could vary across the experimental conditions.

The Problem

Little is known about what observable behaviors are associated with lying. Even less is known about the extent to which observer sensitivity to any behavioral differences between liars and "truthers" allows accurate discrimination between the two. Paul Ekman and Wallace Friesen recently began investigating these questions.

Ekman and Friesen (1969a) develop a theory of the relationship between the nonverbal information "sending capacity" of parts of the body and nonverbal clues to deception. The face, they argue, has a much greater sending capacity than other parts of the body because the facial muscles allow for a larger number of discriminable stimuli patterns. The feet and legs, on the other hand, are the worst information senders, the number of discernible stimuli patterns which can be emitted being far less. The hands are intermediate between the face and feet or legs. The degree

of visibility that the parts of the body have to others in social interaction, say Ekman and Friesen (1969), correspond to their sending capacity. The face is typically more visible during interaction than the hands which in turn are more visible than the feet or legs. "External feedback," they argue, parallels the sending capacity and degree of visibility of various parts of the body. They define external feedback as responses by other people to the focal individual's nonverbal behavior. Most external feedback occurs in response to the face. Individuals are more willing to hold others responsible for, and comment on, information which comes from the face. Less external feedback is directed at the hands and even less at the feet. "Internal feedback," that is, the individual's conscious awareness of his nonverbal behaviors from various parts of the body, also parallels sending capacity, visibility, and external feedback. "People have the greatest internal feedback about their face, next most about their hands and least about their legs and feet" (Ekman and Friesen, 1969a, p. 96).

(Since people receive less external feedback about information which is conveyed from the feet and legs and since they monitor the behavior of their feet and legs less closely, Ekman and Friesen (1969a) reason that the feet and legs will be the greatest source of deception clues, with the hands next, and the face least likely to provide information on which judgments of deception may be accurately based.) Simply put, people are more aware of their face than they are

of the rest of their body. They are thus better able to control their facial behavior during lying. Based on this reasoning, Ekman and Friesen (1969a) hypothesize that observers should be able to detect deceptive communication more accurately when they observe the body than when they observe the head.

Ekman and Friesen (1974) tested this hypothesis by showing videotapes of subjects who were lying or telling the truth in response to questions about their feelings as they watched pleasant and stressful films. Observers were shown videotapes of the subjects with either a close-up shot of the head only or a full shot of the head and body with the head portion blocked from view. Audio portions of the subjects' behavior were not included on these tapes. Results supported the prediction that observers will be more accurate in detecting deception when they view the body only than when they view the head only. Ekman and Friesen (1974) took the most obvious first step in terms of partialling categories of information on which judgments of deception may be based. Their study, however, has left many questions unanswered.

What categories of information provide the best cues on which to judge veracity when the lying involves a factual event rather than the concealment of emotional responses? The Ekman and Friesen (1974) experiment deals only with subjects' false statements about their personal feelings as they watch stressful or pleasant films. Cues which were emitted by subjects may have been in response to these films themselves

and not a result of their false verbal statements about their feelings. It is unknown if the Ekman and Friesen (1974) head-body finding would replicate in other kinds of dishonest interactions.

How accurate would observers be if they were provided with both the head and body for observations? If the head had been available in Ekman and Friesen's "body only" condition, it is likely that observers would have watched it at least some of the time, since it is natural to look at someone's head when they are talking. This, in turn, might have distracted viewers from paying attention to deception cues which were being emitted by the body. In "real life" deception situations the head is obviously available for observation. It is unknown what categories of information observers use to accurately infer deception when the head is available.

To what extent does verbal and paralinguistic information provide cues on which accurate judgments of veracity are based? The audio portion of the lying and truthful behavior of subjects in the Ekman and Friesen (1974) experiment was omitted. The audio portion contains both the verbal content of what is said and paralinguistic cues, which are also audible. Several researchers (e.g., Mehrabian, 1971; Knapp, Hart, and Dennis, 1974) have conducted content analyses of the behavior of lying and truthing individuals and report differences in the verbal content of their speech. Differences between the paralinguistic speech characteristics of lying and truthing individuals have also been demonstrated (e.g.,

English, 1926; Mehrabian, 1971). Research which examines the behavioral correlates of lying does not directly indicate what behaviors observers are using to make accurate inferences of veracity. Differences between the behavior of liars and "truthers" are useful to the observer attempting to detect lying only if the behavior is noticed and correctly attributed to the veracity of the source. Conceivably there could be many observer "cues" indicative of deception which were either missed entirely by observers, or noticed and not correctly identified as being associated with lying. Research which investigates behavioral correlates of lying, while identifying potential categories of behavioral data from which accurate inferences about veracity are made, does not directly provide information about what behavioral cues actually are used by observers. In short, it is unknown how accurate observers could be in detecting deception if they were provided with the verbal content of lying and truthful persons. It is also unknown how accurate they could be if they were provided both the verbal content and the paralinguistic characteristics of subjects' behavior, as would be the case if they were exposed to deceptive and truthful behavior in an audio only format. More importantly, it is unknown how these categories of information would interact with visual information in allowing accurate inferences of veracity. Knapp et al. (1974) suggest that inconsistencies between various categories of information may provide the best cues for making accurate inferences about lying.

In most deception situations, the observer has the opportunity to both see and hear the suspected deceiver. The only way to determine what cues observers actually use to make accurate judgments of veracity is to provide categories of information to observers both in isolation and in combination. The experiment reported in this paper was designed to do exactly this. Its purposes are to refine and extend the work of Ekman and Friesen by: (1) presenting observers with additional categories of information, both in isolation and in combinations, and to compare the resulting accuracy differences; and (2) to examine observer accuracy in a situation involving false statements about a factual event, in addition to replicating the situation used by Ekman and Friesen (1974). It is necessarily "exploratory" because theories of the detection of deception are insufficiently developed and research findings sufficiently scarce to make predictions about accuracy based on these additional categories of information possible. The deception literature does, however, suggest specific cues on which accurate judgments may be based. These cues fall within three categories of "observable behavior": verbal, visual, and paralinguistic. These categories, as well as other relevant concepts, will not be defined.

Some Definitions

It will be useful from the outset to define some important terms and to delimit explicitly the areas of

concern in this paper. Deceptive communication, or lying, as used here, means the purposeful transfer of symbols with the intent to create beliefs in the receiver which are counter to fact as perceived by the source/deceiver. Unintentional lies (misstatements), or lies which are a result of misperceptions of reality by the source as agreed upon by "objective" observers, are thus not within our interest.

The factors which contribute to the accurate identification of deception in everyday encounters are many. Interest here is on observable behaviors which are perceived by the receiver. Situational or context variables will not be considered. This is not to deny their importance. Circumstances such as the receiver's knowledge of the past truthfulness of the source or perceptions of benefits to be accrued by the source from lying probably contribute to the detection of deception. Suspicion of lying by the source may even be a prerequisite for receiver identification of lying.[1] For this reason observer suspicion will be assumed as a constant. However, a thorough examination of the role of situation in detecting deception is beyond the scope of this paper.

The phrase "observable behavior" is broad and potentially ambiguous. Observable behavior, as used here, refers to two categories of acts: verbal and nonverbal. Verbal behavior simply refers to the content or meaning which the linguistic symbols themselves elicit in the receiver. Nonverbal behavior is more difficult to define and the particular areas of interest within a broad class of acts falling

under the rubric "nonverbal communication" need to be explicitly identified.

Researchers have created a variety of nonverbal communication category systems. Knapp (1972) attempts to synthesize various systems and comes up with seven categories.[2] Of those, research examining the detection of deception from observable behaviors has focused on two: body motion and paralanguage. These two categories of nonverbal communication will also be the primary concern in this paper.

Ekman and Friesen (1969b) suggest a five category system for analyzing body movement: (1) emblems, which have direct verbal meaning--e.g., the gestures used to represent "A-OK" or the extended thumb hitch-hiking symbol; (2) illustrators, which serve to illustrate what is said verbally--e.g., pointing when giving directions or depicting spacial arrangements with the hands; (3) affect displays, which are primarily facial expressions of emotion--e.g., anger or fear; (4) regulators, which maintain and control the back-and-forth flow of speaking--e.g., looking up at the other person when it is her or his turn to speak; (5) adapters, which are movements containing fragments of behavior that once had an instrumental purpose--e.g., wiping off one's mouth during conversation or indicating anxiety by rapid hand movements. This seems to be a catch-all category and includes most movement which does not fit clearly into any of the other categories. Body movement provides information to observers through the visual channel only, and the term "visual cues"

will be used to refer to body movements on which judgments of deception are based.

"Paralanguage deals with how something is said, not what is said. It deals with the range of nonverbal vocal cues surrounding common speech behavior" (Knapp, 1972, p. 7). Paralinguistic cues are vocal cues which are audible to the receiver but which by themselves do not have linguistic meaning. Paralanguage consists of "voice set" and "nonverbal vocalizations" (Eisenberg and Smith, 1971). "Voice set" includes such characteristics of the voice such as volume, pitch, rate, and rhythm. "Nonverbal vocalizations" include pauses, audible yawns, belches, and sounds such as "uh-huh," "ah," and "mmm."

Taken together, verbal cues, visual cues, and paralinguistic cues constitute the definition of "observable behavior" to be used in this paper.

One final distinction: The term "subject" will be used to denote the individual who is attempting deception and "observer" will refer to the individual who attempts to detect the deception.

The Deception Literature

Researchers who have studied deceptive communication have approached this topic from two perspectives. One line of research has attempted to identify behavioral correlates of lying by examining differences in the observable communication behaviors of lying and truthing individuals (English,

1926; Berrien and Huntington, 1943; Ekman and Friesen, 1969a; Matarazzo, Wiens, Jackson and Manaugh, 1970; Mehrabian, 1971; Horvath, 1973; Knapp, Hart and Dennis, 1974; McClintock and Hunt, 1975).

A second line of research has examined the extent to which untrained observers can accurately detect deceptive communication (Fay and Middleton, 1941; Hildreth, 1953; Maier and Thurber, 1968; Shulman, 1973; Ekman and Friesen, 1974). Both of these research approaches require the creation of samples of lying and truthful behavior. Most of the methods commonly used to create lying and truthful stimuli have weaknesses which affect interpretation of the results the particular method yields.

Creating Stimulus Materials

A common method used to generate stimulus materials has been to have individuals either pro or counterattitudinally advocate. Knapp et al., (1974), for example, in the most comprehensive content analysis of lying and truthful behavior conducted to date, used this method. Counterattitudinal advocacy simply involves advocating a position counter to one's attitude and Knapp et al. (1974) use this situation to generate their sample of deceptive behavior and employed pro-attitudinal advocacy to get their sample of truthful behavior. This method suffers from three shortcomings. First, if the advocate is simply stating the advantages of a position with which he disagrees, he may not be lying. Since it is quite

possible to hold one position and recognize some advantages in a counter position, the advocate may sincerely believe that the advantages are real. Second, there is evidence that when individuals counterattitudinally advocate they sometimes change their attitude to conform more closely with their verbal behavior (cf. Miller and Burgoon, 1973). Although it is unclear when the "change" takes place, it may conceivably occur before the end of the speech, thus making at least some of the deception stimuli material proattitudinal and truthful. Some researchers suggest that the effects of counterattitudinal advocacy can best be conceptualized as attitude formation rather than attitude change. Bem and McConnell (1970) demonstrate that unless an advocate's attitude is explicitly called to his or her attention and made salient before s/he advocates, s/he may be unable to accurately recall his or her initial position. The final shortcoming of this method is that when individuals counterattitudinally advocate, they may purposely generate observable cues to "tip-off" their audience to the fact that they do not "really" believe what they are saying. These problems make the relationship between the observable behaviors of individuals who are counterattitudinally advocating and those who are lying problematic. While studies using this method to generate stimulus materials may be suggestive of areas to examine, confidence in the results must be qualified.

A second method has been to have subjects role play lying or truthful behavior. Maier and Thurber (1968), in a

study examining the accuracy of observers in detecting deception, used this method. When role-playing, the "lying" behavior is not inconsistent with any matters of fact known to the subject; rather, the subject pretends that he is lying. In this situation, subjects may generate behavioral cues which they believe to be consistent with the behavior of a person who was actually lying in that situation. Whether subjects "know" how real life liars behave is unknown. Role playing as a research method has been severely attacked for this reason (cf. Freedman, 1969). Mehrabian (1972) content analyzed the behavior of individuals who either engaged in counterattitudinal advocacy or role-play lying and found several differences. For example, he reports less frequent head nodding by role-playing liars. At best, it is unclear how the behavior of role-playing liars relates to actual liars and this is important in evaluating the results of studies using this method to create stimulus materials.

A third method used by researchers to generate lying and truthful stimuli has been to have subjects lie or tell the truth in situations where detection has no consequences for the subject. Berrien and Huntington (1941), for example, planted dimes in a classroom and told subjects where to find them. The subject was then asked whether s/he had taken the dime from the room. Subjects were told that if they were detected they would have to give the dime back. Research using instruments to detect deception provides ample evidence that subjects' perceptions of the consequences of being

detected affect the likelihood of detection (Gustafson and Orne, 1963). Davis (1961) suggests a "punishment theory" which states that the greater the consequences of being detected, the greater the physiological response during lying and therefore the greater the likelihood of detection. If the theory also holds for observable behaviors, liars and truthers should behave differently only when there were perceived consequences resulting from detection. The more serious the consequences, the more the behaviors should differ. Several studies which used trivial lies with no consequences resulting from detection, however, did find differences between liars and truthers. An alternative explanation for these differences is the demand characteristics of the research situation. If the subjects are aware that deception is being studied and that there are no consequences if they are detected, they might help observers by purposely generating cues which they believe to be indicative of deception. The applicability of findings based on lying in situations where detection has trivial consequences to situations involving more serious lies is unknown.

A final group of studies has generated stimulus materials in situations where the lies are both real and important. Greatest confidence may be placed in the results of these studies. To properly evaluate the results of studies examining deceptive behavior, it is important to note the method used in creating the stimulus presented to observers.

Figure 1 classifies each study to be reviewed according to the method used.[3]

Research Findings

If a category of information provided to observers is to have an effect on the accuracy of their judgments, behavior available for observation in that category must co-vary in some way with veracity. Findings from studies which examine differences between the behavior of liars and truthers will now be presented within the framework of the three categories of observable behavior defined above.

This review will isolate generalizations (if any are possible) about differential behavior of liars and truthers with respect to that category of observable behavior. It should again be noted that this type of research identifies potential cues only. Differences between the behavior of liars and truthers within a particular category do not necessarily mean that observers will be more accurate when they are provided with this category for observation. To identify veracity they need to both notice the cue and correctly associate it with either lying or truthing.

Behavioral correlates of lying: visual cues.

Head--The Ekman and Friesen (1969a) theory of leakage and clues to deception indicates that the head will be the area of the body least likely to emit cues from which lying can be accurately inferred. Research on head behavior,

Table 1

Method Used to Create Lying and Truthful Stimuli Materials in All Studies Reviewed

Counterattitudinal Advocacy	Role Playing
Hildreth (1953)	Maier (1965)
Mehrabian (1971)	Maier and Janzen (1967)
Experiment I	Maier and Thurber (1968)
Experiment II*	
Knapp, Hart and Dennis (1974)	
Trivial or No Perceived Consequences Resulting for the Subject from Detection	**Serious Consequences**
English (1920)	Ekman and Friesen (1969)
Marston (1920)	Mehrabian (1971)
Fay and Middleton (1941)	Experiment III
Berrien and Huntington (1943)	Shulman (1973)
Matarazzo, Wiens, Jackson and Manough (1970)	Ekman and Friesen (1974)
Cutrow, Parks, Lucas, and Thomas (1972)	
Motley (1974)	
McClintock and Hunt (1975)	
Mehrabian (1971)	
Experiment II	

*Mehrabian (1971, Experiment II) manipulated the method used to induce lying and truthful stimuli including both role playing and counterattitudinal advocacy.

however, suggests there may be some differences between lying and truthful subjects.

The most common area of the head used as the dependent variable in research on deception has been the eyes. McClintock and Hunt (1975) had subjects both truthfully answer questions about pleasant, unpleasant and neutral topics, and lie in answer to questions on a topic of high importance to the subject. There were no consequences for being detected. They found more eye contact when subjects were lying than when they were truthfully answering questions about a pleasant topic. There were no differences in eye contact between the deception condition and conditions in which subjects answered questions about either unpleasant or neutral topics. Knapp et al. (1974) found the duration of eye contact units to be longer when subjects were proattitudinally rather than counterattitudinally advocating. The number of separate eye contact units did not differ in the two conditions. Matarazzo et al. (1970) had subjects lie and tell the truth during an interview. Subjects rehearsed their lies ahead of time and there were no perceived consequences resulting from detection. No relationship between eye contact with the interviewer and veracity was found. These studies provide no basis for generalizations about eye contact and veracity. One study found more eye contact for liars (McClintock and Hunt, 1975), one found more for truthers (Knapp et al., 1974), and one study found no relationship (Matarrazo, et al. 1970).

None of these studies used realistic lying situations with important perceived consequences resulting from detection.

Two other kinds of eye behavior have been examined: pupillary change and blinking. Berrian and Huntington (1943) examined eye pupil change during lying and truthful behavior. They found sudden changes in pupil stability more often when subjects were lying. Berrian and Huntington (1943) used a mechanical "short-focus telescope" to measure pupil change. Since they do not indicate how sensitive this device was relative to what could be observed without a mechanical device, it is unclear if pupil change during lying could be noticed with the naked eye. Obviously an observer would at least have to be in close proximity to notice pupil change. Cutrow et al. (1972) examined eye blinks during deception. They created the perception of moderately serious consequences resulting from detection by telling subjects that "persons of superior intelligence and maturity" could usually avoid detection and that if their lies were not detected they would receive "extra participation credit." They found fewer eye blinks when subjects were lying than when they were telling the truth. Eye blink latency, defined as the length of time between a question and the first blink, did not differ as a function of veracity. Cutrow et al. (1972) used a mechanical device to measure blinking. Observers would also have to be very close to subjects to notice eye blinks.

Several studies have examined smiles and overall facial affect display as dependent variables in deception contexts. Mehrabian (1971) conducted three experiments to examine the behavior of deceitful communicators and used "facial pleasantness" as one of his dependent variables. Experiments I and II used counter and proattitudinal advocacy to generate lying and truthful behavior for analysis and there were minor perceived consequences resulting from detection. Experiment II also had role played lying and truthful conditions. In Experiment III a lying situation was created which probably had important perceived consequences for the subjects. They were induced by a confederate to cheat on a test and were later questioned about how they had done so well. Mehrabian (1971, Experiment I) found higher overall "facial pleasantness" for lying males than for truthful males. No differences were found for females. In Experiment II, subjects who were counterattitudinally advocating displayed greater overall pleasantness than did subjects who were proattitudinally advocating. No consistent pattern was found in Experiment III.

McClintock and Hunt (1975) and Knapp et al. (1974) both investigated the smiling behavior of liars and truthers. Smiles would be subsumed under Mehrabian's (1971) category, "facial pleasantness." McClintock and Hunt (1975) found lying subjects smiled less than truthful subjects, regardless of the pleasantness of the topic about which they were being truthful. Knapp et al. (1974) found no difference on smiles between pro

and counterattitudinal advocates although they do report a slight trend consistent with Mehrabian's findings; i.e., counterattitudinal advocates smiled more than proattitudinal advocates. These findings provide insufficient justification for generalizations about differences between liars and truthers with respect to smiling and overall facial pleasantness.

Hands and Arms--According to Ekman and Friesen (1969a) the hands and arms should emit more cues indicative of deception than the face. Unfortunately, inconsistencies in the operationalizations used to examine hand and arm movement in deception contexts make comparisons among the studies difficult.

Ekman and Friesen (1972) report some preliminary analyses of the stimuli used in their 1974 study. This study probably created quite high perceived consequences of detection by telling subjects that success in their chosen profession was related to the ability to lie successfully. They found that subjects used fewer hand illustrators when lying than when truthing. Liars used more hand shrug emblems and more face play self-adapters than truthers. Knapp et al. (1974) lumped both illustrators and emblems under "gestural duration" and found no difference between liars and truthers. However, this finding may have been consistent with the Ekman and Friesen (1972) result if the latter authors had not distinguished between the two forms of hand movement. Knapp et al. (1974) included Ekman and Friesen's (1972) "self-adapter"

category under a broader category, "adapter duration." Both studies report more of these behaviors in deception conditions, but since Knapp et al.'s (1974) category is broader, it is unknown if differences in the face play self-adapter account for the consistency in the findings, or some other movement within their broader category "adapter duration."

McClintock and Hunt (1975) found no differences in "gestures" which presumably includes both hand shrug emblems and illustrators. They did find more "self-manipulation" in their deception condition than in any of the truthful conditions. "Self-manipulations" as operationalized by McClintock and Hunt (1975) would be broader than, but would include, Ekman and Friesen's (1972) "face-play self-adapter." It would also include other adapter movements in which the hand and arms came into contact with other parts of the body. "Self-manipulation" would be contained in the Knapp et al. category "gestural duration." The issue is further complicated by the fact that all three studies used different methods to create lying and truthful stimuli. Mehrabian (1971) was unable to find a consistent relationship between "self-manipulations" and veracity. Nevertheless, the results of the Ekman and Friesen (1972), McClintock and Hunt (1975), and Knapp et al. (1974) studies might justify a tentative generalization that liars use self-adapters more than truthers.

Legs and Feet--Ekman and Friesen (1969a) indicate that more leakage cues should be emitted by the legs and feet than from other parts of the body. Knapp et al. found no

differences between liars and truthers with respect to leg movements. Mehrabian (1971, Experiment I) found more foot movement by truthers than by liars. He found no differences for leg movement, and in Experiments II and III found no differences for either foot or leg movements.

Several studies have examined posture shifts. Ekman and Friesen (1974) offer a preliminary finding that liars engage in more posture shifts than do truthers. McClintock and Hunt (1975) support this result, finding more posture shifts in the deception condition than when subjects truthfully answered questions about neutral or pleasant topics. However, these studies probably provide insufficient basis for a generalization that the body will be a greater source of leakage cues than will the head. Findings from research which examines the paralinguistic characteristics of speech in deception contexts will now be presented.

Behavioral correlates of lying: paralinguistic cues.

Voice set--Only two studies have examined voice set in deception settings. Mehrabian (1971) measured speech volume of lying and truthful speakers and found no differences in any of his experiments. Motley (1974), using a Kay sonograph instrument which displays sound spectograms, was able to correctly identify veracity in 24 of 36 subjects. Observers, unaided by such an instrument, were unable to exceed chance levels of accuracy. There were no perceived consequences for the subject resulting from detection.

Mehrabian (1971) found speaking rate to be faster for liars than for truthers in Experiment III but not in Experiments I or II. Knapp et al. (1974) did not calculate rate but a rate estimate is obtainable by dividing the total words used by message duration. This yields a slight and probably non-significant trend the opposite of the Mehrabian (1971) result. That is, truthers talked at a slightly higher rate than liars. Other voice set variables such as pitch, range, rhythm, and resonance, have not been investigated.

Nonverbal vocalizations--Mehrabian (1971, Experiment I) found more speech errors in the deceitful condition than in the truthful condition, but this finding was not replicated in Experiments II or III. Knapp et al. (1974) report a non-significant trend in the same direction. Neither study makes clear how speech errors were measured.

Knapp et al. (1974) found no differences in pauses used by truthers and liars. It is unclear, however, how pauses are measured. Matarazzo et al. (1970) found that lying and truthful subjects were equally likely to interrupt the person who was interviewing them.

Four studies have investigated "reaction time latency" and deception. Reaction time latency is the time lapse from when a question is asked to the beginning of the answer. Early researchers (Marston, 1920) thought that liars would require more time to respond because they had to think longer about the answer. Marston (1920) conducted a word association experiment in which subjects were instructed

either to follow directions or to respond in an opposite manner. No consistent relationship between lying and reaction latency was found. Goldstein (1923) found that liars took longer to begin responding to questions than did truthers. English (1926) failed to replicate this finding. All three studies had no perceived consequences resulting from detection. Matarazzo (1970) found truthful and deceitful subjects responded to questions with equal speed. Based on these studies, no generalizations about the paralinguistic characteristics of truthers and liars are justified.

Behavioral correlates of lying: verbal cues.

Knapp et al. (1974) undertook a detailed content analysis of the verbal behavior of liars and truthers. They analyzed twenty categories of verbal behavior and report differences on ten of them at beyond the .01 level (one-tailed). These categories, and the direction of the relationship of liars vs. truthers, are: confidence ratio, liars less than truthers; different words, liars less; factual statements, liars less; self-experience, liars less; past references, liars less; probes, liars more; self-interest, liars less; other references, liars, more; disparaging statements, liars, more; and total words, liars less. Message length was also measured by Matarazzo et al. (1970) and Mehrabian (1971), but operationalized as time duration instead of total words. Mehrabian (1971) in Experiment II found role playing male truthers talked longer than role playing liars. The

difference was not found for the pro and counterattitudinal advocacy conditions of the experiment or for Experiments II or III.

Since none of these studies created stimuli in situations with definite perceived consequences for the liar resulting from detection, there is no firm basis for generalizations about the relationship between veracity and verbal content. Research which has focused on the detection of deception from observable cues will now be examined.

Detecting deception. The second major approach used by social scientists to study deceptive communication has been to examine the extent to which observers, unaided by mechanical devices, can accurately detect lying. These studies provide information about the particular behavioral differences that discriminate truthers and liars only if the categories of behavior which are presented to observers are manipulated. If observers are given identical information, it is possible to make statements about their accuracy in detecting deception, but it is not known on what basis the judgments were made.

Fay and Middleton (1941) had subjects answer questions about their personal characteristics over a public address system. There were no perceived consequences for being detected. They found correct judgments of veracity in 55.6% of the trials. Hildreth (1953) found that observers were unable to distinguish between pro and counterattitudinal advocates at beyond a chance level. He, too, did not create

conditions of perceived consequences resulting from detection for his subjects. Maier (1965) found that role playing liars and truthers were accurately detected at a better than chance level. Again there were no consequences for detection. Shulman (1973) created a realistic lying situation modeled after Mehrabian's (1971, Experiment III) technique in which subjects were induced to cheat on a task and then asked how they had done so well. His observers were unable to detect deception at beyond chance levels.

Maier and Janzen (1967) attempted to identify the specific cues that lead to accurate judgments by asking observers to list the observable behaviors which resulted in particular judgments of veracity. They found no relationship between accuracy and the reasons cited for judgments. Observers may not have known the basis for their judgments and thus may have generated a list of post hoc reasons to rationalize their decision about any given individual.

Two studies have manipulated the categories of lying and truthful behavior made available to observers and compared accuracy scores. Maier and Thurber (1968) had subjects role play lying and truthing behavior in an interview situation. Subjects' answers to questions were provided to observers in three different formats. One group watched the interviews live (watchers). A second group heard a tape recording of the interviews (listeners) and the final group read a written transcript of the stimuli material (readers). Results indicated that listeners and readers were more

accurate (77.0% and 77.3%, respectively) than watchers (58.3%). The authors reason that the visual cues available to the watchers may have distracted them from noticing the verbal (and paralinguistic) cues and thus reduced their accuracy scores.

Ekman and Friesen (1974) conducted the only experiment which manipulated the categories of behavior presented to observers and created lying and truthful stimuli in a situation with important perceived consequences resulting from detection. Student nurses were told that success in lying was associated with professional success. They were shown silent films which were either pleasant or unpleasant and asked to describe their feelings while watching them. Immediately before the film began they were instructed to answer either honestly or untruthfully. Observers were shown silent videotapes of these answers which contained either a close-up shot of the head, or a shot of the head and body with the head blocked from view. Half of the observers first viewed a truthful sample of each subject's behavior. In the conditions where no truthful "familiarity" segment was initially presented, subjects who viewed the body only were no more accurate than those who viewed the head only. However, when the familiarity segments were included, subjects were more accurate when they saw the body only than when theysaw the head only (56.8% vs. 45.4%).

The Ekman and Friesen procedures for generating honest and deceptive behavior do have one shortcoming. The

unpleasant films they used were extremely stressful. The cues which they emitted may have indicated extreme emotional stress not directly associated with lying. Specifically, the cues could have been indicative of the subjects' failure to control their emotional response to the films themselves, and not a result of their verbal statements about their feelings. The behaviors which subjects emitted could have occurred even if they were not questioned about their feelings as they watched the films. The experiment reported here attempts to improve external validity by having subjects answer questions about a factual event, in addition to answering questions about their feelings while viewing stressful materials.

Summary and Research Design

The research on visual, paralinguistic, and verbal correlates of lying and truthing behavior offers little in terms of identifying specific cues on which accurate judgments of deception may be based. However, each of these categories include at least several studies which do report differences between the behavior of liars and truthers. Visual, paralinguistic, and verbal behavior all contain the potential for accurate inferences of veracity.

The studies on detecting deception demonstrate that untrained observers, unaided by mechanical devices, are able to detect deception more accurately in some circumstances than in others. The studies by Maier and Thurber (1968) and

Ekman and Friesen (1974) suffer from procedural weaknesses that limit their contribution to our understanding of the behavioral basis for accurate inferences of deception. However, both studies suggest useful approaches to the problem and demonstrate that it is possible to manipulate categories of behavior presented to observers and to observe systematic differences in accuracy. The present study attempts to refine and extend the work of these researchers by presenting observers with categories of observable behavior in isolation, and in combinations, and comparing the resulting accuracy of judgments of veracity.

Visual information will be operationalized in this experiment as "shot," which represents physical areas of the bodies of subjects made available to observers, and "colorbw," which represents whether observers are exposed to the stimuli in a color or a black and white format. The shot variable will have three levels: head only; body only; and head and body. Head only and body only conditions will allow the direct replication of Ekman and Friesen (1974) while the head and body condition will permit an examination of accuracy in a situation more closely approximating the visual information available to observers in "real life" deception contexts.

The colorbw variable has two levels: color, and black and white. The Ekman and Friesen (1974) experiment used black and white videotapes, yet visual stimuli are perceived in color in face-to-face situations. Certain potential

visual correlates of deception, such as facial flush or perspiration, may be noticed more readily in color than in black and white. Including this format variable also increases the potential applicability of the results. For example, considerable recent discussion has dealt with applications of video tape technology in the legal system (Bermant and Jacoubovitch, 1975; Doret, 1974; McCrystal, 1972; Morrill, 1971), and the use of videotape in court proceedings is increasing. Accurate detection of lying is important for equitable juror decisions. The type of television format (color or black and white) could affect jurors' ability to detect lying. Thus, including this variable creates the possibility of generating findings with important applications in the court system.

Observer accuracy based on paralinguistic information will be examined in several ways. A transcript only condition, with all paralinguistic information removed, will operationally define the verbal content of the subject's responses. Comparisons between this condition and an audio only condition will allow an examination of the extent to which paralinguistic cues alone contribute to accuracy. Observers in the Ekman and Friesen (1974) experiment did not hear the audio portion of the tapes. Half of the observers in this experiment, who base their judgments solely on visual information, will also not have the audio information available. The other half will both see and hear the behavior of the subjects.

Those conditions which include visual information (i.e., all conditions except audio-only and transcript-only) create a 3 (shot: head-only/body-only/head-and-body) by 2 (colorbw: color/black and white) by 2 (visaudio: visual-only/visual-and-audio) factorial design. The audio-only and transcript-only conditions bring the total number of cells in the design to 14. The design is graphically displayed in Figure 1.

We will now turn to a discussion of how and why observer accuracy in identifying subject veracity could vary across levels of each independent variable.

Shot

Head only vs. body only. If Ekman and Friesen (1969a) are correct in theorizing that the body provides the greatest source of leakage and clues to deception, their finding that observers are more accurate when judgments of veracity are based on the body-only rather than on the head-only should be replicated. Their 1969 paper makes a distinction between deception clues and leakage. Deception clues tip off an observer that deception is in progress but do not reveal the nature of the concealed information. Leakage indicates what information is being withheld. Leakage would thus only be available in situations where affective responses were being concealed. If the content of the verbal statements were about complex factual events, rather than merely statements about emotional feelings, it would be impossible for nonverbal

	Head Only		Body Only		Head and Body	
	Audio and Visual	Visual Only	Audio and Visual	Visual Only	Audio and Visual	Visual Only
Color						
Black and White						

Audio Only

Transcript Only

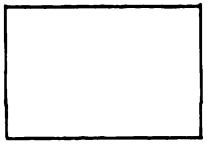

Figure 1. The Experimental Design.

cues to convey accurately the nature of the withheld information. Thus, leakage would not be possible and judgments of veracity would have to be based on deception clues. If, as has been suggested above, the Ekman and Friesen (1974) head-body finding was a result of subjects' body movements caused by their general discomfort as they watched stressful films, the body may not be the best source of information about lying when the content of the false statements is factual rather than emotional. In short, the original finding could have been the result of leakage, i.e., the failure of subjects to control their emotional feelings. Thus, while we might expect to replicate Ekman and Friesen's head-and-body finding with lying and truthful stimulus materials similar to theirs, it is less clear what will happen when the stimuli are based on statements about concrete matters of fact. Consequently no directional prediction will be offered about the accuracy of observers who view the head-only versus those who view the body-only.

Head-and-body vs. head-only and/or body-only. Observers who view the head-and-body obviously have more available information than those who view either the head-only or the body-only. Whether the head or the body is the greatest source of cues indicative of deception, observers in the head and body conditions have those cues available, and they should have the greatest chance to identify veracity accurately. Also if, as Knapp et al. have suggested,

inconsistencies between categories of behavior result in the identification of veracity, observers in the head-and-body condition would have the best opportunity to notice such "out of sync" behaviors between the head and body. For these reasons, they could be more accurate.

On the other hand, if the greater information sending capacity of the head results in more deception cues emitted from this area, the head-and-body condition could result in less accuracy, because less head detail would be available for observers in the head-and-body condition than for those in the head-only condition. Also, the head might distract viewers from noticing body cues or vice-versa and less accuracy might result. For these reasons, observers in the head-and-body conditions could be less accurate than those in either the body-only or head-only conditions. Since arguments can be marshaled in favor of each condition resulting in higher observer accuracy scores, and since previous research provides no clear-cut basis for a directional prediction, none will be offered.

Colorbw

Research on the effects of viewing color versus black and white television formats indicates there may be greater information retention of peripheral material when the presentation is in color (Katzman, 1971; Katzman and Nyenhuis, 1971). As was pointed out above, several potential visual correlates of deception, such as facial flush and perspiration, may be

noticed more readily in color than in black and white, particularly in the head-only condition of the shot variable. This greater richness and variety of color cues could result in higher observer accuracy when the stimuli are viewed in color. Also color may increase observer interest in the stimuli, causing them to pay closer attention and to be more sensitive to noticing cues indicative of deception.

On the other hand, a color presentation might distract observers from noticing cues indicative of deception. For example, brightly colored clothing might receive more observer attention in a color format than in a monochromatic format. Since there is no clear-cut basis for a directional prediction, none will be offered.

Visaudio (Paralinguistic)

Observers in the transcript-only condition will make their judgments based on the verbal content of the subjects' behavior. In addition to this same verbal content, audio-only observers will also have paralinguistic information available. Such paralinguistic information could provide cues indicative of deception which, in turn, could result in greater accuracy. However, paralinguistic cues could also distract observers from scrutinizing carefully the verbal content for any inconsistencies which would allow the accurate identification of veracity.

A similar situation exists in comparing the visual-only with the visual and audio conditions. The audio band

could provide cues from which accurate judgments could be made. Visual and audio observers would also have the opportunity to notice any inconsistencies between the verbal and nonverbal behavior of subjects which were associated with deception. The audio band, on the other hand, could distract observers from noticing visual cues signaling deception, or vice-versa. Here too, then, as with the other variables, there is no solid foundation on which to base directional predictions and none are offered. Thus, this experiment is an exploration of how the verbal and nonverbal behaviors available to observers influence their ability to detect deception.

Footnotes

[1]Shulman (1973) tested this proposition and failed to find any support for the notion that forewarning the observers that lying might occur increased the accuracy of judgments of veracity.

[2]Knapp's entire category system for organizing various kinds of nonverbal communication includes physical characteristics, touching behavior, proxemics, artifacts, and environmental factors, in addition to body movement and paralanguage to be defined below.

[3]The consequences of detection for the subject is, of course, only one side of the coin. Observers' motivation to accurately judge veracity probably also affects accuracy. Unfortunately most of the studies provide no information about observer motivation and consequently none will be presented here.

Chapter II

METHOD

Overview

Samples of both lying and truthing communication behavior were videotaped under conditions of high saliency for 23 subjects. These tapes were edited to include eight subjects who were making true or false statements about the factual content of an event they had seen and eight subjects who were making true or false statements about their emotional feelings as they watched pleasant and stressful slides. Two versions of each tape were made, one of which was the exact inverse of the other; that is, if a particular subject were shown lying on one tape, that same subject was always shown truthing on the other and vice-versa. One or the other of these two tapes was shown to 730 college student observers under 14 different conditions thus creating a total of 28 cells in the design. Observers saw either a close-up of the subjects' head-only, a shot of the body-only, or a shot of the head-and-body. They viewed these tapes in either a color or a black and white format. Half of the observers viewed the tapes without the audio channel present and half were allowed to hear the audio. Half the observers viewed one

version of the tape and half viewed the other. These variables created a 3 (head-only/body-only/head-and-body) by 2 (color/black and white) by 2 (visual-only/audio-and-visual) by 2 (Tape I/Tape II) factorial design. Audio-only and transcript-only conditions (for both Tapes I and II) brought the total number of cells in the design to 28. Measures of observer confidence in their judgments were also obtained.

Procedurally, then, this research was conducted in two phases. First the stimulus materials, which consisted of videotapes of subjects engaging in lying and truthful behavior, were created. These videotapes were then shown to groups of observers under each of the manipulated experimental conditions. Presentation of the research methods in this chapter follows this same logical order.

Creating the Stimulus Tapes

Saliency

It was important to create a situation in which successful lying was difficult. If the behavior of lying subjects were no different than when they were telling the truth it would be impossible to interpret observer accuracy scores across the experimental conditions. Put another way, if there were no cues which were indicative of deception, observers would probably not be able to differentiate between lying and truthful subjects at different levels of accuracy in the various conditions; and even if they did, it would be impossible to ascertain why. Thus, many of the procedures

used to create the stimulus materials were designed so that it would be difficult for subjects to lie without being detected.

Ekman and Friesen (1969a) discuss the conditions under which lying will be maximally difficult. First the deception should be salient fcr both the source (subject) and receiver (observer). "Saliency" refers to the extent to which the deceiver is consciously concerned about the deception, and to the extent to which it is important to the deceiver to avoid detection. The deceiver should be highly ego-involved with his or her success in deceiving the observer. Under these conditions, Ekman and Friesen (1969a) argue, cues giving away the deception are most likely to occur. This argument is supported by Davis' (1961) theory discussed earlier and Gustafson and Orne's (1963) finding that the greater the perceived consequences resulting from detection, the greater the physiological response during lying. Deception will also be difficult, Ekman and Friesen (1969a) suggest, when the subject is in the role of both deceiver and detector while the observer is only in the role of detector. In this situation the subject wants to both deceive the observer and monitor the behavior of the observer to determine the success or his or her deception.

An effort was made to create both roles for the subjects. Saliency was created in a manner similar to that used by Ekman and Friesen (1974). Subjects were 19 male and four female senior criminal justice majors at Michigan State

University, all of whom planned careers in law enforcement. They were sent a letter by the Director of the School of Criminal Justice requesting that they participate in a research project designed to "identify certain personal characteristics of individuals which may contribute to their successful performances as police officers" (see Appendix A). When the subjects arrived to participate in the study they were told that the research was attempting to develop improved screening tests for prospective police officers; that earlier research had shown that those who could lie successfully were better officers than those who could not; that the Criminal Justice Department was very interested in how well they, in particular, did on this task; and that feedback about their performance would be provided the department and could affect such things as their letters of recommendation for securing positions as police officers. They were questioned during the taping by a detective from the East Lansing Police Department whose purpose was to add credibility to this cover story. In an effort to place the subject in the dual role of detector and deceiver, the subjects were also told to monitor the officers' reactions to determine if they were successfully deceiving him.

Stimulus Content

The interviews, during which the stimulus tapes were made, consisted of four segments. First, subjects were asked five questions about personal characteristics, all of

which they answered truthfully. As was noted earlier, Ekman and Friesen (1974) were able to support their head vs. body accuracy prediction only when they provided the observers with a sample of honest behavior first. It may be necessary to have some familiarity with the behavior of an individual to be able to accurately identify lying. The truthful segment was provided to increase the likelihood that observers would be able to detect lying and thus the manipulated categories of information would be more likely to produce systematic differences.

The second segment involved questions about the factual content of a videotape subjects had seen before being interviewed. They were shown one of two versions of this videotape. Both versions showed the sentencing of a criminal who had been tried and found guilty of murder. The individual who was being sentenced reacted on one tape very violently. He swung at his attorney, attacked the prosecutor and bailiff, and had to be forcibly removed from the courtroom. In the other version of this tape, the individual was very passive in hearing his sentence. He listened to the judge politely and was quite docile as he was led from the courtroom. Subjects were instructed to lie to the first three of the five questions they were asked about the content of this videotape and to tell the truth to the last two questions, or vice versa. Subjects who were instructed to lie were also told what to say, so as to cut down on the variability of the content of the lying responses. If subjects had been able to

make up their own lies, the content of those who were telling the truth would have been very consistent across subjects, while the content of those who were lying would have had great variability. The variability factor probably would spuriously increase observer accuracy.

The last of the five questions during this second segment was open ended and simply asked for a description of the defendant's reaction to being sentenced. Subjects typically took considerably longer to answer this question than the other four. On this question, the answer which lying subjects were instructed to give was consistent with the version of the tape which they did not see. In other words, those who saw the violent version were told to say that the defendant responded in a very docile manner while those who saw the nonviolent version were told to say that he reacted violently. Thus, the same answer could be true for one subject and false for the other. To further cut down on the variability of the answers to this fifth question, subjects were given a slip of paper on which phrases were written summarizing the reaction of the defendant on the version of the tape which they did not see. This was designed to help them recall the form of their lying response. This second segment of the interview is henceforth labelled the "factual" segment of lying or truthful behavior.

The last two segments of the interviews consisted of questions about subjects' feelings or emotions as they watched color slides on a television screen in front of them.

One sequence of slides was very pleasant and showed scenes with boats, lakes, children playing, and so on. The unpleasant sequence of slides showed explicit close-up pictures of people who had been seriously burned. Subjects were told that immediately before each sequence of slides began, they would see a slide with either the word "lie" or "true" on it. If it said "true" they were to answer truthfully about their feelings. If it said "lie" they were to respond with answers which were the opposite of their feelings. They were falsely told that some subjects would see two sequences of pleasant slides, some two sequences of unpleasant slides, and some one of each. They did not know which sequence would be prefaced by the word "true" and which with "lie." In reality, the word "true" always prefaced the pleasant sequence, and the word "lie" the unpleasant sequence. Thus the content of their answers was always that they were experiencing pleasant feelings, regardless of which slides they were viewing.[1] The order in which the sequence of slides was shown was randomized across subjects. These segments of the interview will subsequently be labelled the "emotional" segments.

In summary, tapes were made of each subject: (1) telling the truth in answer to questions about their personal characteristics; (2) lying and telling the truth about the factual content of the videotape they had seen; (3) lying about their feelings as they watched unpleasant slides; (4)

telling the truth about their feelings as they watched pleasant slides.

Procedures

Subjects were contacted by phone two to three days after receiving the letter from the Director of the School of Criminal Justice requesting that they participate in the research. The caller identified himself as being associated with the Department of Communication and asked the subject if he or she had received the letter. The caller then briefly repeated the content of the letter and asked the subject if s/he could come in for one hour to participate several days later.

Thirty-five potential subjects were sent letters. Of these, 28 were successfully reached by phone. Twenty-six individuals agreed to participate. They were scheduled to arrive every half hour from two in the afternoon running into the evening hours over a two-day span.

When a subject arrived s/he was greeted by the author (E_1) and thanked for coming. The following is an accurate paraphrase of the cover story which was given them:

> Hello, I'm John Hocking and I'm with the Department of Communication. Thank you for coming. I know that Mr. Brandsteader was very anxious to have you participate.
>
> This research is being funded by the RANN division of the National Science Foundation. RANN stands for "research applied to national needs." This project has been going on at several universities and police agencies for about three years now. Essentially this research is an extension of some earlier research

which was conducted by the New York police training and performance group and research which was done by the RANN corporation on police background characteristics and performance. As the letter from Mr. Brandsteader indicated, we're interested in identifying certain personal characteristics of individuals which may contribute to performance as successful members of the law enforcement profession.

As I'm sure you're aware, there has been considerable dissatisfaction with present techniques for screening prospective police officers. Presently paper and pencil tests are the primary means of screening applicants. Some individuals turn out to be poor officers even though they score well on the tests. Others, who might make good officers, are in some cases eliminated from consideration by these paper and pencil tests. The research you'll be participating in today is designed to identify specific behaviors which could be used as predictors of successful job performance. Tests on these behaviors could then be used to supplement the traditional paper and pencil tests now used.

Your background records were reviewed by the criminal justice department and they selected you for participation in this study. What are your career goals? (Pause for response.) Ah, I can see why the department is so interested in your performance on this research.

One of the consistent findings of the earlier research in this area is that successful members of the law enforcement profession are able to lie successfully under a variety of circumstances. For example, when interrogating someone, it might be necessary to give the impression that you know more than you do. There are other examples, too. When you're dealing with people under stress circumstances, it's pretty important to respond to them in an appropriate way, even if you feel differently. You might need to appear calm and in control when you don't feel that way at all. In short, members of the law enforcement profession need to be pretty good actors at times.

Well, the research that we're doing here today involves giving you the opportunity to attempt to successfully lie under several different circumstances. The earlier research has shown that individuals who later proved to do well in their jobs as police officers have been extremely successful at this task. I guess you can see why Mr. Brandsteader was so anxious to find out how certain people in his department performed in this research.

Subjects were then told that they would be interviewed by a police officer from the East Lansing Police Department and the format of the interview was explained. After being shown the tape on which the factual segment of the interview was to be based, they were instructed about which questions to answer truthfully and untruthfully. Subjects were also given the information on which to base their lying responses. They were told the general information the factual questions would be seeking, but were not told specifically how the questions would be worded. After they understood the format of the factual questions they were told about the emotional segments, and were given the opportunity to decline to view the burn victims' slides with no penalty. All subjects agreed to view those slides. Any questions they had were answered in a manner consistent with the cover story. After E_1 was sure each subject understood what was required of him/her, he took the subject to the nearby television studio and introduced him/her to the police detective. During the walk to the studio, it was explained rather vaguely that the answers would be videotaped for later analysis. The prebriefing took an average of about 20 minutes.

The detective seated the subject and then sat down in his own chair which was about 15 feet directly in front of the subject. He then reinforced each of the major points which had been made by E_1 during the prebriefing cover story, indicating that there was dissatisfaction with current screening tests for prospective officers and that research

had indicated that good liars made good officers. He described how in his own work a frequent strategy used during interrogation was to exaggerate the amount of information he had about a crime in order to elicit information from the suspect. Only to the extent that he was successful in convincing the suspect that he really had the information was this a good strategy. The officer explained that he was involved in the research because the East Lansing Police Department was interested in adopting the new screening procedure. The last thing the officer told the subjects before beginning the questioning was that they should do their very best because the E.L.P.D. and the Criminal Justice Department were both "very interested" in their performance.

While the officer was talking with the subject, $\underline{E}_1$ gave a form to $\underline{E}_2$ in the control room of the television studio which summarized the instructions the subject had been given about when to lie and when to tell the truth during the factual questions. This form also indicated which sequence of slides the subject was to view first during the emotional questioning. As the interview was conducted, $\underline{E}_2$ monitored the subjects' answers and noted any deviations from the instructions (see Appendix B).

Subjects were videotaped with two color cameras. One had a close-up shot of the subject's head while the other had a full shot of the head and body, including the feet. Consequently, less head detail was observable in the head-and-body shot than in the head-only shot. The head-and-body

shot also showed the microphone and microphone stand. The background for both camera shots was a light blue curtain which was about eight feet (2.5 meters) behind the subject, thus appearing slightly out of focus. While subjects wore colors and styles of clothing which varied considerably, none were dressed in an unusually loud or gaudy manner. The subject was seated in a moderately comfortable chair. Two Sony 8600, half-inch, reel-to-reel videotape recorders were used to record the subjects' answers. A professional video technician ran all recording equipment and adjusted the close-up camera for each subject. He also controlled the film chain operating in the control room through which the slides were shown to the subject on a television monitor. Both cameras were about 25 feet away from the subject and sitting next to each other, thus keeping the angle of the two shots relatively constant. The close-up picture was achieved with a telephoto lens. This camera placement resulted in camera shot angles from slightly to the left of where the officer was sitting. Thus in looking at the officer as they answered questions, subjects were looking slightly to the left of the cameras. The television monitor on which the slides appeared during the emotional segments was sitting on an eight-inch high platform in front of the officer's desk and slightly to his right. The subject thus had to look to the officer's left and slightly down to watch the slides. The monitor was 12-13 feet (about four meters) from the subject. The officer could not see the monitor from where he was sitting and

consequently from the subject's point of view (and in reality) he did not know which slides were being shown when. Figure 2 graphically displays this room arrangement. Subjects were then asked the 22 questions in the following order.

Truthful Sequence

1. What is your name?
2. What year in school are you?
3. What are you majoring in?
4. Where are you from?
5. Did you happen to go to the football game with Ohio State a couple of weeks ago? What did you think?

Factual Sequence

6. What was the crime which Mr. Bostick was convicted of?
7. Did Mr. Bostick say he was sorry for his crime?
8. Did the judge say that Mr. Bostick has the right to appeal his conviction?
9. What was Mr. Bostick wearing?
10. What was Mr. Bostick's reaction to being told that all he needed to do to appeal his conviction was to sign a form?

Slide Sequence 1

11. What kinds of feelings are you having right now?
12. What kind of mood do these slides create?
13. What other experiences have you had which convey the same feelings as these slides?
14. What are your feelings now that the slides are over?

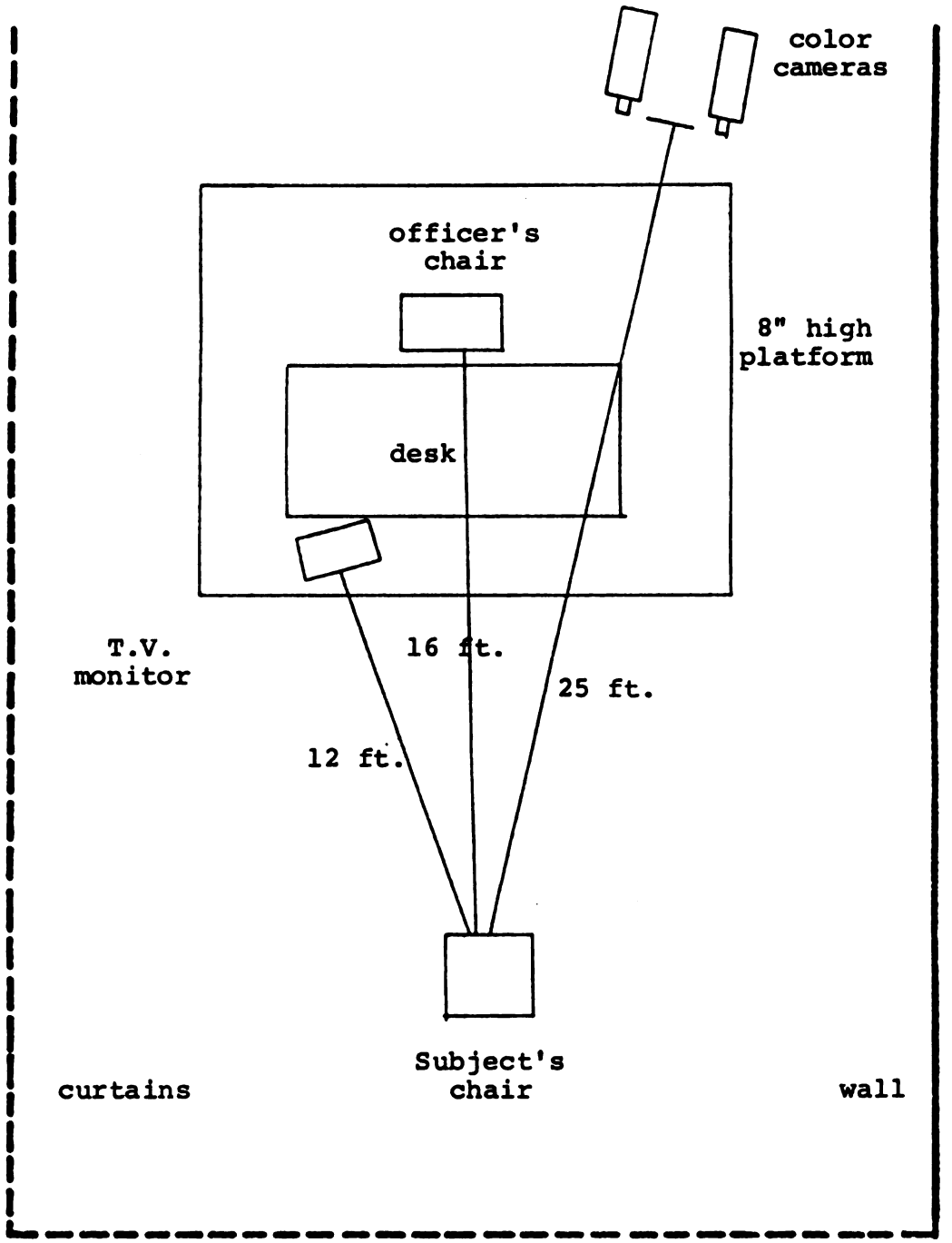

Figure 2. Room arrangement during the stimulus taping

15. Are you telling me the truth?

16. Do you think I believe you?

Slide Sequence 2

17. What kinds of feelings are you having right now?

18. What kind of mood do these slides create?

19. What other experiences have you had which convey the same feelings as these slides?

20. What are your feelings now that the slides are over?

21. Are you telling me the truth?

22. Do you think I believe you?

The entire interview sequence took about five to seven minutes. After the interview was over, the subject was thanked by the police officer and introduced to E_3 who debriefed the subject.

Debriefing

Subjects were taken to a nearby snack bar by E_3. Before beginning the debriefing, they were given a short one-page questionnaire to fill out (see Appendix C). The primary purpose of this questionnaire was to find out how the subject actually felt when s/he watched the slides of burn victims. It was possible that some subjects might have said they felt pleasant when watching these slides and been telling the truth. An examination of responses to an open ended question indicated that all subjects selected for inclusion on the emotional segments of the stimulus tapes did indicate feeling unpleasant while watching the burn slides. This

questionnaire also asked subjects how successful they felt they had been at deceiving the officer during both of their lying sequences.

During the debriefing, subjects were told that the Criminal Justice Department would not be receiving information about their performance (see Appendix D). They were asked for their permission to show their tape to undergraduates at Michigan State the following term and to a group of adults from the Lansing area. They were asked not to discuss the research with other criminal justice majors who might be participating, until after the second day of running. They were told that they would soon be learning the full details of the research.

Editing the Tapes

Two subjects failed to show up for participation, one subject refused to participate after the prebriefing had been completed, and the tapes of two subjects were lost because of equipment failure. Taping procedures were completed for 19 male and four female subjects. This resulted in approximately four hours of lying and truthful behavior on tape--two hours of the head shot and two hours of the head-body shot. These tapes had to be edited into master stimulus tapes of no longer than 25-30 minutes to make data collection in a 50 minute class possible. It would have been desirable to use longer segments of each subject and to use samples of all 23 subjects. This would have necessitated collecting data

outside of class. The large number of conditions necssitated a very large number of observers (see page 60 below). Unfortunately it would have been impossible to obtain this many observers outside of class.

Whatever the observable cues from which inferences of veracity are made, it is safe to assume that not everyone emits them equally. Fay and Middleton (1941), for example, found that subjects were judged accurately by as few as 50.9% of the observers and by as many as 62.8%. Some subjects in the present experiment probably emitted many cues which would likely be perceived by observers as indicative of lying both while they were lying and truthing. If by chance a lying segment were selected from such persons and included on a tape for judgment, observers' accuracy scores would probably be spuriously inflated. This would not be because they were able to discriminate between the lying and truthing behavior of the persons, but rather because they always looked like they were lying. Other people may not give off many revealing cues; these persons would probably look like they were telling the truth all the time. If a truthful segment of these persons' behavior were included on a tape, accuracy scores would also be increased; once again not because of the observers' ability to detect veracity, but because of an idiosyncrasy of those people's behavior. Analogously, if opposite segment of behavior were selected, observer accuracy scores would probably be spuriously deflated. An effort was made to solve this problem by creating two tapes, one of which was the

inverse of the other. Thus if a truthful segment was selected at random for inclusion on Tape 1, a lying segment from the same person would be included on Tape 2, and vice-versa. This procedure resulted in an additional experimental manipulation thus doubling the number of conditions in the design from 14 to 28.

Whether a segment of any given subject's behavior was included from the factual or the emotional portion of their interview was assigned at random except in cases where the subject's failure to follow instructions resulted in only the factual or the emotional segment being available. Eight truthful and eight lying segments were assigned at random to Tape 1. Of these, four of each were factual and four of each were emotional. Tape 2 was created by taking the opposite behavior from the segment (factual or emotional) for each subject.

There were several subjects who failed to provide both a lying and a truthful segment for either the factual or emotional portions of the interview. One of these individuals was selected for inclusion as the practice example (see below) for observers, and the rest were not included on the tapes. This left a total of 16 subjects, 13 male and three female.

The format of each master stimulus tape was identical. Each subject was shown twice. The truthful segment appeared first and was followed by the "test" segment which was either truthful or not. Before each of these segments, an announcement was made on all tapes which told observers

which segment they were about to see (truthful or test) and which subject was about to appear. After the test segment, the announcer stated, "That was the test segment for subject number 1 (or 2, . . . , or 16), please mark your questionnaires." There was then a 10 second pause before the truthful segment for the next subject was announced.

Before the 16 subjects who were to be judged appeared on the tape, a practice example was presented in the same format as the rest of the tape. It was, however, labelled the "practice example."

The physical editing was done by a professional video technician. Editing was accomplished by recording from the original tape onto master stimulus tapes in the proper sequence. The announcements which accompanied each segment were recorded onto the master tapes at the same time. Since it was necessary that the head only tapes have exactly the same segments of the subjects' behavior as the head and body tapes (and vice-versa), this was accomplished by making note of the verbal content of the audio portion of the beginning and end of each segment. It was thus possible to record this segment of behavior onto the master tape made for the other camera angle by closely monitoring the verbal content of the tape.

Four tapes which included the audio portions of the questions and answers were made: head only (Tape 1), head and body (Tape 1), head only (Tape 2), head and body (Tape 2). The final number of master stimulus tapes was doubled by the

tapes which contained the visual behavior only. These were created by making a copy of each of the above four tapes with the audio recording jack removed during the segments of the subjects' behavior. Thus, the tapes to be used in the visual-only conditions were identical to those in the video and audio conditions, except for the absence of the audio. The announcements of segments and subject numbers were on all tapes.

All tapes were in color. In the black and white conditions they were played by using black and white television monitors. Tape 1 was 27 minutes and 15 seconds long, while Tape 2 was 26 minutes and 20 seconds long.

The transcript condition for both tape versions was made from the audio conditions. It included the verbal answers only. All paralinguistic cues such as 'ahs," "mmm," and other audible sounds which do have corresponding verbal symbols were removed. This was done so that the comparisons between the transcript-only conditions and the conditions which included audio would partial out only the verbal content.

The audio-only condition was created by playing the audio and visual tapes with the video portion of the television monitors adjusted so that there was a black screen. Thus the sound quality was identical to the sound which was played in the video and audio conditions.

Collecting the Data

Observers

The observers who judged the veracity of the subjects appearing on the stimulus tapes were 730 undergraduate male and female students enrolled in 18 introductory communication classes during the spring term of 1975 at Michigan State University. The two classrooms in which all classes met each had room dividers which allowed the class to be broken into two groups. This was done by having students with even student numbers on one side of the divider and those with odd student numbers on the other. This usually resulted in nearly equal numbers of observers in both groups. On those occasions when this was not the case, observers were arbitrarily moved to create equal sized groups. Thus there were 36 separate intact groups which were randomly assigned to the 28 experimental conditions. The only constraint on this random assignment was that two conditions which contained the audio band were always run at the same time, because the room dividers were not completely sound-proof. The 28 largest groups were each assigned to one condition. The remaining eight smallest groups were run in the same condition as the eight smallest groups within the 28. This was done to create as large and evenly distributed sample size across conditions as possible. Table 2 presents the sample size obtained in each of the 28 cells.

Table 2

Sample Size in Each Cell of the Design

	Tape	Head Only: Audio and Visual	Head Only: Visual Only	Body Only: Audio and Visual	Body Only: Visual Only	Head & Body: Audio and Visual	Head & Body: Visual Only		
Color	I	25	39	30	27	33	25	179	331
	II	26	26	31	22	23	24	152	
Black and White	I	28	18	21	25	24	27	153	301
	II	24	24	27	22	28	23	148	
		103	107	109	96	108	109		
		210		205		217			

	Audio Only	Transcript Only
Tape I	23	19
Tape II	24	21

Audio and visual = 320
Visual only = 312

Questionnaire

The questionnaire on which observers indicated their judgments (see Appendix E) had three pages of instructions which explained that the research was examining people's ability to detect lying and described in some detail the circumstances under which the stimulus tapes had been created. For example, it was explained that the first eight people were answering questions about a videotape they had seen and the last eight were answering questions about slides that they were viewing. Observers were told that it was very important to each of the subjects that they successfully deceive the police officer who was interviewing them. The presentational format for the various segments on the tape was also explained. To motivate the observers to pay close attention to the tapes, they were told that if they put their name and summer mailing address in the place provided, they would be sent their personal accuracy score. It was emphasized that participation was voluntary and they should do so only if they were willing to do their very best to detect whether or not a subject was lying.

An effort was made to avoid observers' attempts to make an equal number of judgments of truthing and lying by explicitly telling them that the particular tape which they saw might contain mostly truthful segments, mostly lying segments, or about equal numbers of each. Observers were told to make each judgment independent of their other judgments. They were also told there was no relationship

between the length of a segment and whether the subject in that segment was telling the truth or lying, and that the same answers could be truthful for some subjects and untruthful for others.

Observers made judgments dichotomously for each subject. Following each judgment, observers indicated how confident they were of their judgment on an 11-point scale. There were 16 such sets of measures, corresponding to the 16 test segments on each stimulus tape. Each observer was thus required to make 16 separate judgments of veracity and 16 assessments of their degree of confidence.

The questionnaire also collected information about how successful observers perceived themselves to be at lying, how interested they were in participating in the experiment, and whether they were sitting in a good position to observe the tape. Standard demographic information was also gathered.

Procedures

Two Sony color television monitors were set up (and turned on) on one side of the classroom and two black and white monitors were on the other, prior to the arrival of the students. Both monitors were run off the same videotape recorder and thus showed identical pictures. When the body-only visual conditions were run, the upper third of the monitors were blocked from view with a piece of cardboard.

Several minutes after the class was scheduled to begin the instructor introduced E_1 as a graduate student in the Department of Communication who was conducting some interesting research. E_1 then briefly explained that the research was being funded by the National Science Foundation and was examining individuals' ability to detect lying. It was explained that some students had been interviewed by a police officer the previous fall and had been instructed to lie at certain times and to tell the truth at others. Observers were told that if they chose to participate, they would be watching tapes which were made of these individuals and attempting to detect lying.

After this brief introduction, the observers were divided into two groups as described above, and the room divider was closed. E_2 and an assistant were with one group of observers and E_1 and an assistant were with the other. The questionnaires were then handed out. Each questionnaire had previously been coded by condition to reduce the possibility of subsequent confusion. While the questionnaires were being handed out, those observers who appeared to be in a bad position for viewing the tape were asked to move to a better spot.

The instructions on each questionnaire were read aloud and observers were instructed to follow along. After the instructions were read, the tape was started and the observers watched the practice example. When the practice example was completed the tape was turned off and observers

were asked if they had any questions. When all questions were answered and the Es were satisfied that all observers understood what was required of them, the tape was started. The last thing said before the tape began was a reemphasis of the importance of making independent judgments without consulting with other observers. Students who came into class late after the instructions had been read were allowed to sit in the room during the experiment but did not fill out a questionnaire.

After the tape was completed, subjects were asked to finish filling out the questionnaire. The questionnaires were then collected, the room dividers opened, and the observers debriefed.

The experiment was run over an eight day period, beginning on one Friday and concluding the following Friday (April 11-18, 1975).

Debriefing

In the 50 minute classes, of which there were 11 (out of 18), there was only time for an extremely short debriefing. This consisted of reaffirming that there was no deception used during this phase of the research and explaining the other conditions in the experiment. Observers were assured that they would be receiving the promised information about their accuracy scores if they had requested it and thanked for participation. In the longer classes a more detailed debriefing was presented if the instructor

was willing to allow the additional time. This longer debriefing included showing samples of tapes from other conditions, describing procedures through which the stimulus tapes were made, and a more thorough explanation of the various questions that the research was designed to answer. The Ekman and Friesen head-body hypothesis was not explicitly mentioned for fear of observers talking about the study with other students who were yet to participate. Except for this, no effort was made to keep details of the research secret. There was also no effort to reduce interclass discussions about the experiment.

Results Letter

A letter briefly summarizing the experimental design, the major findings, and individual personal accuracy scores was sent to those observers who had requested this information. This letter was sent about four months after the data were collected (see Appendix F). The importance of the observers' contribution to the research was emphasized and they were again thanked for participation.

Footnotes

[1]It may have been desirable to have had subjects also respond with lying answers as they watched the pleasant slides and truthing answers as they watched the burn slides. Ekman and Friesen (1974), however, only showed segments to observers of subjects truthing during pleasant films and lying during unpleasant films. In order to replicate them as closely as possible, the same procedure was followed here.

CHAPTER III

RESULTS

The primary dependent variable in this experiment was observer accuracy in identifying subject veracity. Each observer made 16 accuracy judgments, eight of subjects engaged in lying or truthing behavior about the factual content of the videotapes they had seen and eight of subjects making true or false statements about their emotional feelings as they watched pleasant and stressful slides. The accuracy scores for each observer were obtained by separately adding the number of correct judgments within both segments and dividing by eight. Thus, each observer's level of accuracy is represented by two proportions: one for the factual segments and one for the emotional segments. Scores consequently have a possible range of 0 to 1.0, with .50 representing four correct judgments out of the eight attempted.[1]

Observers' degree of confidence in their accuracy judgments was also treated separately for the factual and emotional segments. These responses were added and divided by eight thus giving confidence scores a range of 0 to 10. Identical analyses were performed on each of the four

dependent variables: factual accuracy; emotional accuracy; factual confidence; and emotional confidence.[2]

Four-way analyses of variance were performed on the data from the 24 cells which fit within the completely crossed factorial design.[3] Comparisons between the audio-only and transcript-only conditions, and other comparisons, were performed with *t* tests. These tests used the MS_W from the overall analysis when cells included in the comparison were from the overall analysis of variance. The MS_W used when the transcript-only and audio-only conditions were included in comparisons were calculated by averaging the variance within each of these cells into the MS_W from the overall analysis of variance. Any comparison which exceeded the .05 level two-tailed was considered significant.[4]

Tape is not a theoretical variable in this experiment. The purpose of creating two versions of the stimulus was to have both a lying and truthing sample of each subject's behavior judged by observers. It was assumed that observers would make more judgments of "true" (or "false") for some subjects than for others, regardless of the subject's actual veracity. Observer accuracy in judging each subject's veracity across both levels of the tape variable will now be presented, followed by a presentation of the results for both accuracy dependent variables. The chapter concludes with the presentation of the results for observers' degree of confidence in their accuracy judgments.

Tape

Table 3 presents the number of judgments of "true" and "false" made of each subject on each tape, summing across all 14 conditions. This table also displays the mean accuracy and confidence scores for each subject on each tape, as well as the summed scores from both tapes. Examination of this table shows that some subjects were judged to be truthing by a majority of observers both when they were truthing and when they were lying, and vice versa. For example, subject 13 was judged to be lying by 74.3% of the observers when he was truthing and by 69.1% when he was actually lying. Accuracy was thus much higher when he was lying (.691) than when he was truthing (.257). Overall, the accuracy with which this subject was judged was .474, which probably represents a better accuracy estimate than if only a truthful or a lying sample of behavior had been included on the stimulus tape. Further examination of Table 3 reveals similar examples for many of the 16 subjects.

Table 4 presents the means and variances for factual accuracy in all cells; Table 5 presents the analysis of variance for the 24 cells which fit within the factorial design; and Tables 6 and 7 present this information for emotional accuracy. Examination of Table 5 indicates a significant main effect for Tape ($p < .001$) on factual accuracy. The tape I mean for all 14 cells was .567 while the tape II mean was .512. For factual accuracy, tape also enters into interactions with both visaudio and shot. These interactions will be

Table 3

Subject Sex, Actual Veracity, Number of Judgments of "True" and "False," and Mean Accuracy and Confidence Scores for Both Factual and Emotional Segments on Each Tape, and Summed

Subject			Tape I					Tape II				Both Tapes				
			No. of Judgment					No. of Judgment				Totals				
#	Sex	T-F	False	True	Accuracy Mean	Conf Mean	T-F	False	True	Accuracy Mean	Conf Mean	Accuracy Mean	Conf. Mean	Total False	Total True	True
1	F	F	174	202	.463	5.79	T	121	229	.654	5.49	.559	5.64	295	431	.593
2	M	T	247	131	.347	5.80	F	223	127	.637	5.57	.492	5.68	470	258	.355
3	M	F	274	104	.727	6.01	T	179	171	.489	6.09	.608	6.05	453	275	.378
4	F	T	84	294	.778	6.66	F	183	166	.524	6.04	.651	6.35	267	460	.633
5	F	T	194	184	.487	6.01	F	247	102	.708	6.30	.598	6.16	441	286	.393
6	M	F	268	109	.711	6.20	T	196	154	.440	6.00	.576	6.10	464	263	.362
7	M	T	210	167	.443	5.86	F	92	258	.263	5.64	.353	5.75	302	425	.585
8	M	F	219	158	.581	5.98	T	211	138	.395	6.12	.488	6.05	430	296	.407
Factual Totals			1670	1349	.567	6.04		1452	1345	.514	5.91	.541	5.97	3122	2694	.463
9	M	F	275	103	.728	6.87	T	166	183	.524	6.09	.626	6.48	441	286	.393
10	M	F	140	238	.370	6.67	T	208	141	.404	6.25	.387	6.46	348	379	.521
11	M	F	243	135	.643	6.09	T	190	159	.456	5.92	.550	6.00	433	294	.404
12	M	T	171	207	.548	6.26	F	200	150	.571	6.61	.560	6.44	371	257	.491
13	M	T	280	97	.257	6.55	F	242	108	.691	6.28	.474	6.42	522	205	.282
14	M	T	197	180	.477	6.21	F	104	245	.298	5.57	.388	5.89	301	425	.585
15	M	T	168	208	.553	6.01	F	120	229	.344	5.88	.449	5.95	288	437	.601
16	M	F	241	135	.641	6.28	T	181	169	.483	6.22	.562	5.79	422	304	.418
Emotional Totals			1715	1303	.527	6.37		1411	1384	.471	6.10	.499	6.18	3126	2687	.462

Table 4

Cell Mean Accuracy Scores, Variances, and Marginal Means for the Factual Segments*

			Shot						
			Head Only		Body Only		Head & Body		
	Tape		Audio and Visual	Visual Only	Audio and Visual	Visual Only	Audio and Visual	Visual Only	Marginal means
Color	I	$\bar{x}$	.615	.545	.533	.431	.682	.550	.559
		s^2	.025	.034	.028	.034	.040	.025	
	II	$\bar{x}$	.601	.438	.561	.466	.592	.385	.506
		s^2	.030	.027	.023	.309	.039	.016	
	Marginal means								.533
Black and White	I	$\bar{x}$	.580	.535	.568	.405	.599	.598	.548
		s^2	.048	.035	.041	.024	.037	.033	
	II	$\bar{x}$	.575	.400	.551	.460	.567	.386	.490
		s^2	.040	.017	.033	.017	.034	.034	
	Marginal means		.593	.480	.553	.441	.610	.480	.519
			.537		.497		.545		

Transcript Only

Tape	
I	.625 .070
II	.625 .011

Audio Only

Tape	
I	.671 .018
II	.565 .023

visual & audio = .585
visual only = .467
grand mean = .540

*Note: See Table 2 (page 60) for sample sizes.

Table 5

ANOVA for Factual Accuracy

Source of Variation	Sum of Squares	DF	Mean Square	F	Probability Level
Main Effects	2.889	5	.578	18.527	.001
Colorbw	.040	1	.040	1.296	.254
Visaudio	2.033	1	2.033	65.189	.001
Tape	.563	1	.563	18.068	.001
Shot	.342	2	.171	5.487	.005
2-way Interactions	.914	9	.102	3.258	.001
Colorbw	.029	1	.029	.940	.999
Colorbw tape	.009	1	.009	.275	.999
Colorbw shot	.031	2	.015	.492	.999
Visaudio tape	.169	1	.169	5.428	.019
Visaudio shot	.002	2	.001	.036	.999
Tape shot	.632	2	.316	10.131	.001
3-way Interactions	.322	7	.046	1.473	.173
Colorbw visaudio tape	.001	1	.001	.021	.999
Colorvw visaudio shot	.104	2	.052	1.674	.186
Colorbw tape shot	.003	2	.001	.044	.999
Visaudio tape shot	.219	2	.110	3.519	.029
4-way Interactions	.046	2	.023	.737	.999
Colorbw visaudio tape shot	.046	2	.023	.737	.999
Residual	18.711	600	.031		
Total	22.882	623	.037		

Multiple $\underline{r}^2$ = .126

Table 6

Cell Mean Accuracy Scores, Variances, and Marginal Means for the Emotional Segments*

	Tape		Head Only		Body Only		Head & Body		Marginal means	
			Audio and Visual	Visual Only	Audio and Visual	Visual Only	Audio and Visual	Visual Only		
Color	I	$\bar{x}$	.535	.468	.563	.583	.559	.548	.543	
		s^2	.020	.023	.018	.029	.012	.039		
	II	$\bar{x}$	.481	.486	.491	.455	.462	.422	.466	
		s^2	.020	.032	.020	.019	.030	.026		
										.505
Black and White	I	$\bar{x}$	.513	.417	.542	.600	.521	.537	.522	
		s^2	.021	.026	.021	.020	.023	.016		
	II	$\bar{x}$	.448	.448	.449	.473	.469	.456	.457	
		s^2	.019	.023	.027	.018	.024	.025		
										.490
	Marginal means		.494	.455	.511	.528	.503	.491		
			.475		.520		.497			

Transcript Only

Tape	
I	.522 .025
II	.521 .020

Audio Only

Tape	
I	.421 .013
II	.524 .013

visual & audio = .511
visual only = .491
grand mean = .501

*Note: See Table 2 (page 60) for sample sizes.

Table 7

ANOVA for Emotional Accuracy

Source of Variation	Sum of Squares	DF	Mean Square	F	Probability Level
Main Effects	1.002	5	.200	8.901	.001
Colorbw	.018	1	.018	.802	.999
Visaudio	.020	1	.020	.902	.999
Tape	.743	1	.743	32.995	.001
Shot	.227	2	.114	5.044	.007
2-way Interactions	.328	9	.036	1.618	.106
Colorbw visaudio	.011	1	.011	.476	.999
Colorbw tape	.006	1	.006	.282	.999
Colorbw shot	.027	2	.013	.590	.999
Visaudio tape	.000	1	.000	.005	.999
Visaudio shot	.077	2	.039	1.718	.178
Tape shot	.205	2	.103	4.557	.011
3-way Interactions	.173	7	.025	1.100	.361
Colorbw visaudio tape	.003	1	.003	.152	.999
Colorbw visaudio shot	.023	2	.012	.521	.999
Colorbw tape shot	.027	2	.014	.609	.999
Visaudio tape shot	.119	2	.059	2.636	.071
4-way Interations	.001	2	.000	.020	.999
Colorbw visaudio tape shot	.001	2	.000	.020	.999
Residual	13.511	600	.023		
Total	15.015	623	.024		

Multiple $\underline{r}^2$ = .067

described when the results of these theoretical variables are presented.

Examination of Table 7 indicates a significant emotional accuracy main effect for tape ($p < .001$). The tape I mean for all 14 cells was .524 while this figure for tape II was .470. The tape variable also enters into an interaction with shot. The nature of this interaction will be described in the context of describing the effect of shot. Clearly, then, the rationale for including two versions of the stimulus tape is supported by the results. The results for the primary dependent variable in this experiment, observer accuracy, will now be presented.

Accuracy

Shot

Factual. Table 5 shows a significant main effect for shot ($p < .005$). Inspection of the marginal means in Table 4 indicates that those observers who viewed the body-only had a mean of .497 which was less accurate than both head-only observers, who had a mean of .537 ($t = 2.31$, $df = 413$, $p < .05$), and head-and-body observers who had a mean of .545 ($t = 2.80$, $df = 421$, $p < .02$). Table 5 also indicates a shot by tape interaction. Table 8 presents the cell means which resulted in this interaction while Figure 3 displays the relationship graphically. Examination of Table 8 coupled with subsequent t tests shows that both the main effect for shot and the interaction between shot and tape were a result

Table 8

Cell Means, Sample Sizes, and Marginal Means for the Two-Way Interaction Between Shot and Tape on Factual Accuracy

		Shot			
Tape		Head	Body	Head & Body	Marginals
I	$\bar{x}$	.569	.484	.607	.553
	n	110	103	119	
II	$\bar{x}$	.504	.510	.483	.499
	n	100	102	98	
Marginals		.537	.497	.545	

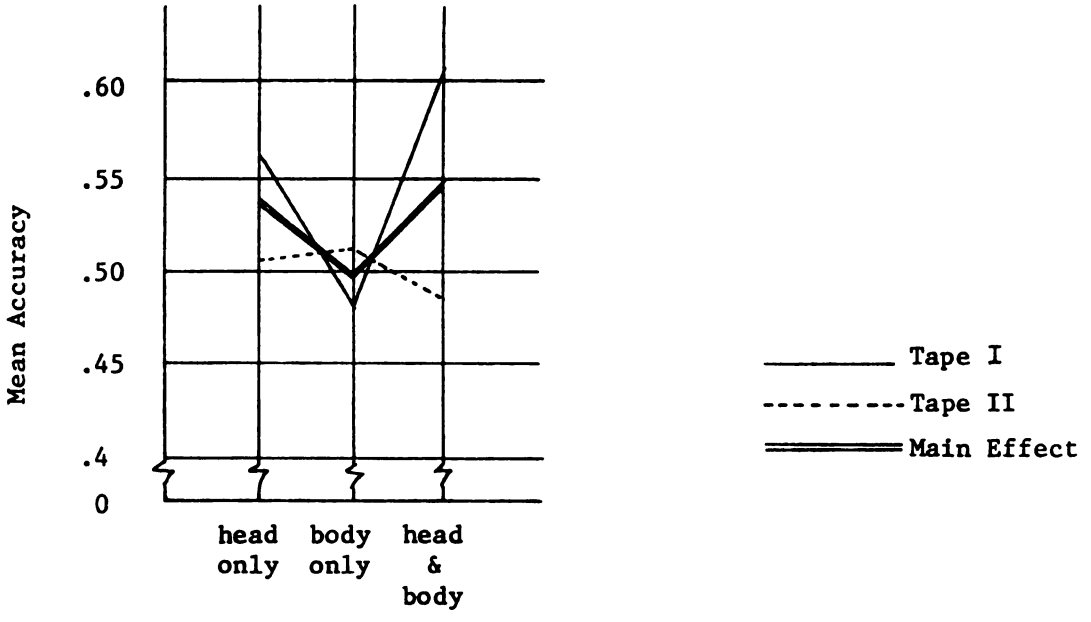

Figure 3. Two-Way Interaction Between Shot and Tape on Factual Accuracy

of the cell mean from tape I. Within tape I, the body-only condition, with a mean of .484, was significantly lower than both the head-only mean of .567 (t = 3.59, df = 218, p < .001) and the head-and-body mean of .607 (t = 5.13, df = 213, p < .001). The head-only condition does not differ significantly from the head-and-body condition (t = 1.58, df = 213).

Within tape II, an examination of Table 8 indicates that the cell mean in the body-only condition is largest with a mean of .510, but this does not differ significantly from either the head-and-body mean of .483 (t = .84, df = 198) or the head-only mean of .504 (t = .24, df = 204).

Table 5 also shows a significant interaction between shot, tape, and visaudio. The cell means resulting in this effect are shown in Table 9, and this interaction is graphically displayed in Figure 4.

Table 9

Cell Means, Sample Sizes, and Marginal Means for the Three-Way Interaction of Visaudio, Shot and Tape on Factual Accuracy

		Shot						
		Head Only		Body Only		Head & Body		
Tape		Aud. & Visual	Visual Only	Aud. & Visual	Visual Only	Aud. & Visual	Visual Only	Marginals
I	$\bar{x}$	.598	.540	.551	.418	.641	.574	.554
	n	53	57	51	52	57	62	
II	$\bar{x}$	.588	.419	.556	.463	.580	.386	.499
	n	50	50	58	44	51	47	
Marginals		.593	.480	.554	.441	.611	.480	.527

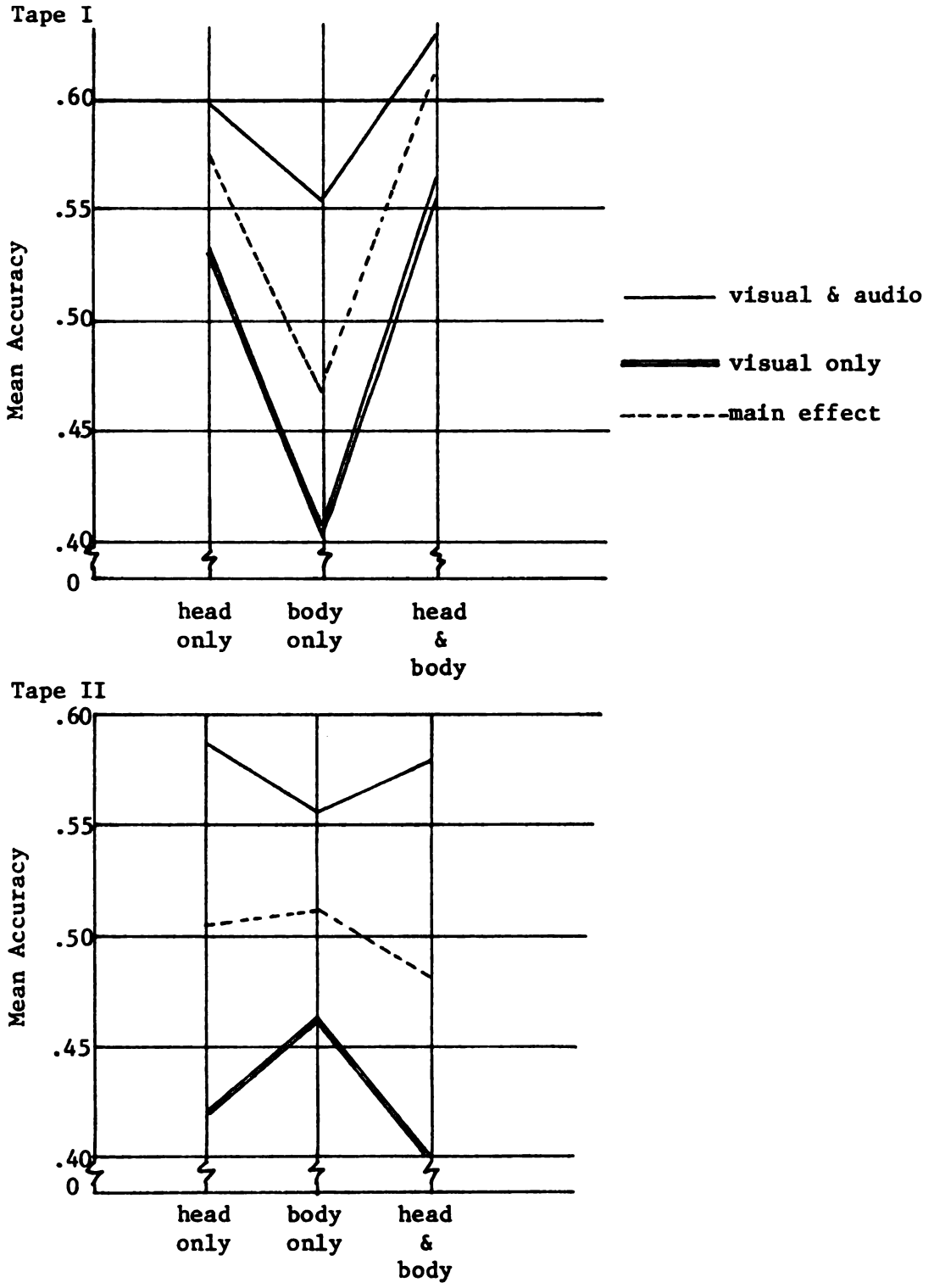

Figure 4. Three-Way Interaction Between Visaudio, Shot and Tape on Factual Accuracy*

*These same relationships are displayed from the visaudio perspective in Figure 7, page 83.

The Ekman and Friesen (1974) experiment dealt with the shot variable in a visual-only context. The head-only visual-only cell mean of .480 did not differ significantly from the body-only mean of .441 (t = 1.58, df = 202).

Emotional. Examination of Table 7 indicates that there is a significant main effect for shot ($p < .007$). Inspection of the marginal means in Table 6 indicates that subjects in the body-only condition, with a mean of .520, were more accurate than subjects who viewed either the head-only (.475) or the head-and-body (.497). Table 7 also indicates a significant shot by tape interaction. Table 10 displays the cell means which resulted in this interaction while Figure 5 displays this relationship graphically. Subsequent t tests showed that the main effect for shot was not general across conditions but was limited to tape I, where the head-only condition mean of .483 was significantly less than both the body-only mean of .572 (t = .431, df = 214, $p < .001$) and the head-and-body mean of .541 (t = 2.81, df = 231, $p < .005$). The body-only mean was not significantly larger than the head-and-body mean (t = 1.52, df = 222). Within tape II the body-only conditions mean of .467 did not differ significantly from either the head-and-body mean of .452 (t = .70, df = 200) or the head-and-body mean of .466 (t = .04, df = 201).

Examining those cells in Table 6 which replicate Ekman and Friesen (1974), the head-only visual-only cells had

Table 10

Cell Means, Sample Sizes, and Marginal Means for the Two-Way Interaction Between Shot and Tape on Emotional Accuracy

		Shot			
Tape		Head	Body	Head & Body	Marginals
I	$\bar{x}$	.483	.572	.541	.532
	n	110	103	119	
II	$\bar{x}$	.466	.467	.452	.462
	n	100	102	98	
marginals		.475	.520	.497	

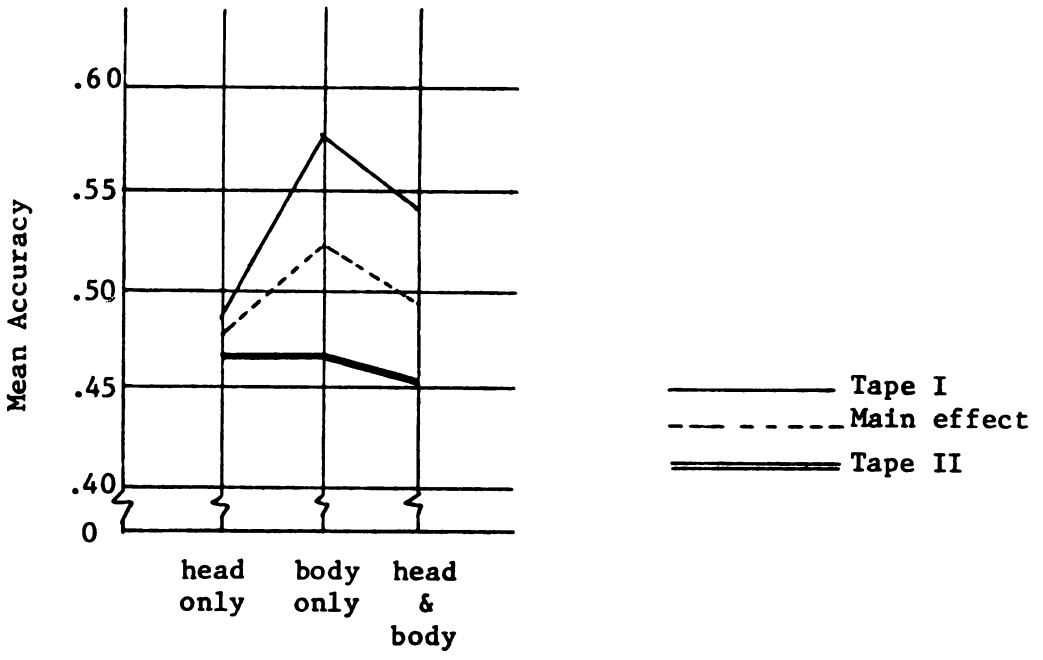

Figure 5. Two-Way Interaction Between Shot and Tape on Emotional Accuracy

a mean of .455 which was significantly smaller than the body-only visual-only mean of .528 (t = 3.42, df = 202, $p < .002$). Also, within the visual-only conditions, the head-and-body mean of .491 does not differ significantly from either the head-only (t = 1.74, df = 216) or the body-only (t = 1.74, df = 206).

Visaudio

Factual. Table 5 indicates that there is a main effect for visaudio ($p < .001$). An Examination of Table 4 indicates that for all 12 pairs of cells, observers who were exposed to both the visual and audio information were more accurate than those whose judgments were based on visual information only. The mean for all 12 audio and visual cells was .585 while the mean for the 12 visual-only cells was .467.

Table 5 also indicates a visaudio by tape interaction. Examination Table 11 and Figure 6, which display this interaction, indicates that the effect of the visaudio variable was more pronounced for tape III. Visaudio also enters into a three-way interaction with shot and tape. Figure 7 graphically displays this interaction from the perspective of the visaudio variable.

Emotional. Examination of Table 7 indicates that for emotional accuracy, no main effects were found for visaudio, nor did this variable enter into interactions.

Table 11

Cell Means, Sample Sizes, and Marginal Means for the Two-Way Interaction Between Visaudio and Tape on Factual Accuracy

Tape		Visual-Audio	Visual Only	Marginals
I	$\bar{x}$	.596	.511	.554
	n	161	171	
II	$\bar{x}$	.575	.423	.499
	n	159	141	
Marginals		.585	.467	

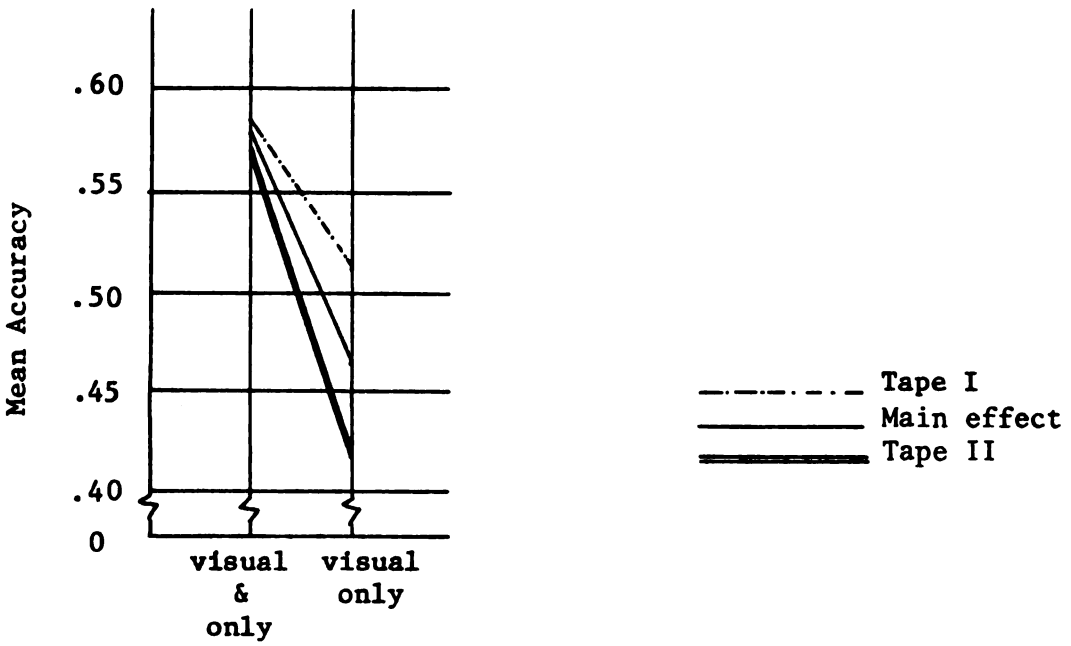

Figure 6. Two-Way Interaction Between Visaudio and Tape on Factual Accuracy

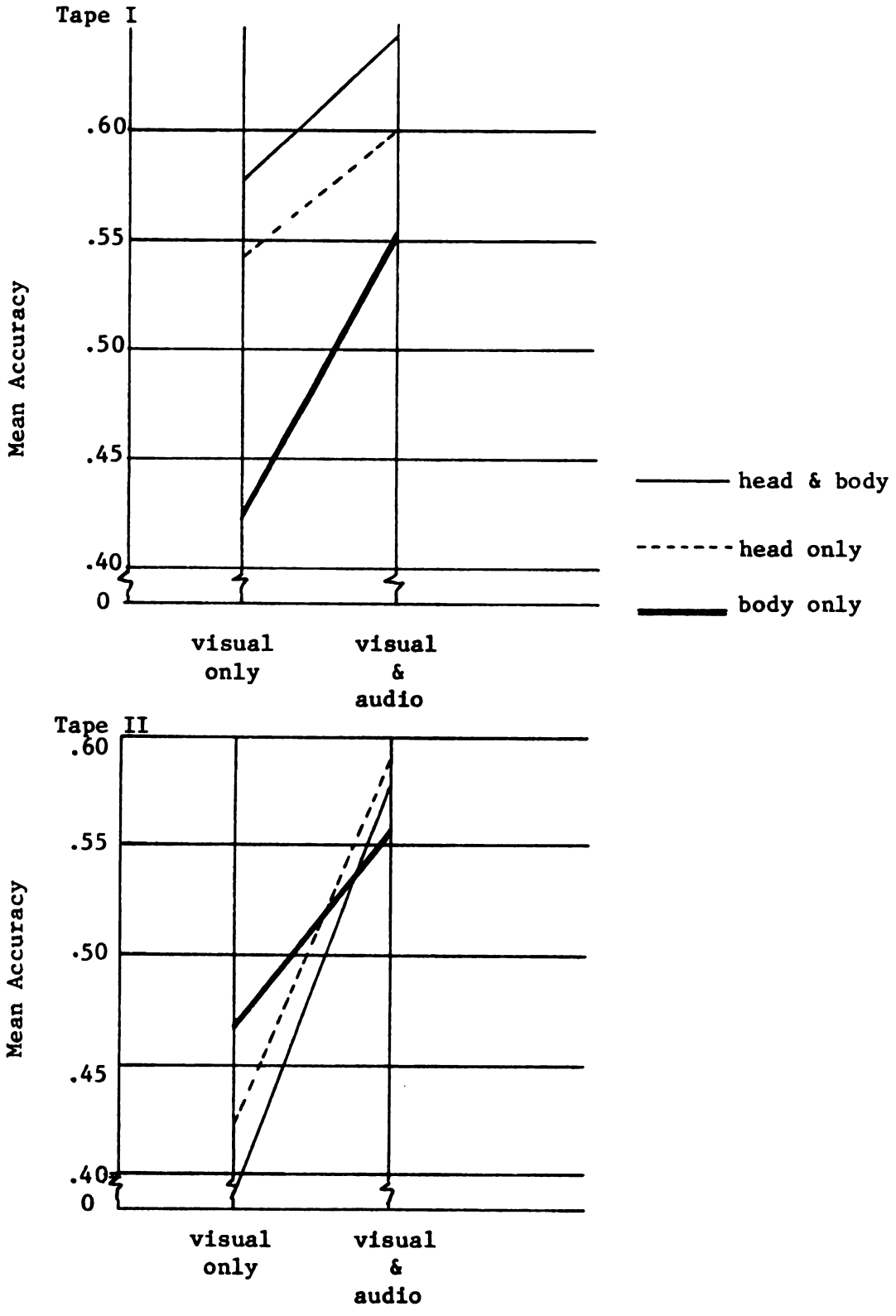

Figure 7. Three-Way Interaction Between Visaudio, Shot, and Tape on Factual Accuracy

Color/Black and White (Colorbw)

Factual. An examination of Table 5 indicates that no main effects or interactions were found for colorbw for factual accuracy.

Emotional. An examination of Table 7 indicates that no main effects or interactions were found for colorbw for emotional accuracy.

Transcript-Only and Audio-Only

Factual. Observers in the transcript-only condition had a mean accuracy score of .625 (for both tape I and II) while observers in the audio-only condition had a mean of .618. This difference is not significant (t = .22, df = 184).

The visual-only mean accuracy score of .467 for all 12 visual cells was significantly less than the audio-only mean of .618 (t = 5.24, df = 351, $p < .001$). The audio-only condition also resulted in higher observer accuracy than all three visual-only levels of the shot variable. Both the head-and-body and the head-only conditions had means of .480 (t = 4.51, df = 146 and 147, respectively, $p < .001$), while the body-only condition had a mean of .441 (t = 5.67, df = 134, $p < .001$).

The visual and audio mean factual accuracy score across all 12 conditions was .585. This value does not differ significantly from the audio-only mean of .618 (t = 1.14, df = 360). Within the visual and audio conditions,

only the body-only condition, with a mean of .553 differed significantly from the audio-only mean of .618 (t = 2.12, df = 148, $p < .05$). The head-only, visual and audio conditions had a mean of .593 (t = .81, df = 141) and the head-and-body, visual and audio conditions had a mean of .610 (t = .26, df = 146).

Emotional. Observers in the transcript-only conditions had a mean accuracy score of .522, while observers in the audio-only conditions had a mean of .473. This difference was not significant (t = 1.71, df = 85).

The visual-only mean accuracy score of .491 for all 12 cells was not significantly different from the audio-only mean of .473. Within the visual-only conditions, only the body mean of .528 was significantly greater than the audio-only mean (t = 2.09, df = 135, $p < .05$). The head-only mean of .455 (t = .69, df = 145) and the head-and-body mean of .491 (t = .70, df = 149) did not differ significantly from the audio-only mean of .473.

The visual-and-audio mean emotional accuracy for all 12 cells was .503, which does not differ significantly from the audio-only mean (t = 1.22, df = 359). All three levels of the shot variable within the visual and audio conditions also failed to differ significantly from the audio-only mean. Head-only had a mean of .494 (t = .81, df = 141), body-only had a mean of .511 (t = 1.47, df = 147), and the head-and-body

condition had a mean of .503 (t = 1.16, df = 147). Results for observer confidence will now be presented.

Confidence

Table 12 presents the means and variances for factual confidence in all cells; Table 13 presents the analysis of variance for the 24 cells which fit within the factorial design; and Tables 14 and 15 present this information for emotional confidence.

Tape

Factual. Examination of Table 13 reveals that, for factual confidence, there were no significant main effects or interpretable interactions[5] for tape.

Emotional. Examination of Table 15 indicates a significant main effect for tape ($p < .025$). Tape I had a mean for all 14 cells of 6.38 while tape II had a mean of 6.12. Tape did not enter into any interpretable interactions with the other variables.

Shot

Factual. An examination of Table 13 reveals that there was no significant main effect for shot, nor does shot enter into interpretable interactions with the other variables. However, a comparison between the head-only mean of 6.13 with the body-only mean of 5.77 yields a significant t of 2.36 (df = 413, $p < .05$).

Table 12

Mean Confidence Scores, Variances, and Marginal Means for the Factual Segments*

			Shot						
			Head Only		Body Only		Head & Body		
	Tape		Audio and Visual	Visual Only	Audio and Visual	Visual Only	Audio and Visual	Visual Only	
Color	I	$\bar{x}$	6.55	5.97	5.87	5.80	6.58	5.56	6.06
		s^2	1.87	1.04	2.91	2.73	1.57	2.67	
	II	$\bar{x}$	6.47	6.00	6.39	4.91	5.96	5.68	5.90
		s^2	3.24	4.46	1.91	2.59	1.97	3.05	
									5.98
Black and White	I	$\bar{x}$	6.15	6.05	6.71	5.11	5.52	6.20	5.96
		s^2	1.98	1.80	2.11	3.46	2.41	2.00	
	II	$\bar{x}$	6.39	5.40	5.63	5.70	6.22	5.70	5.84
		s^2	1.31	3.46	2.20	2.90	2.89	1.79	
									5.90
Marginal means			6.39	5.86	6.15	5.38	6.07	5.79	
			6.13		5.77		5.93		

Transcript Only

Tape	
I	6.52
	2.30
II	5.58
	5.65

Audio Only

Tape	
I	5.92
	1.49
II	6.12
	1.16

audio & visual = 6.20
visual only = 5.68
grand mean = 5.96

*Note: See Table 2 (page 60) for sample sizes.

Table 13

ANOVA for Factual Confidence

Source of Variation	Sum of Squares	DF	Mean Square	F	Probability Level
Main Effects	52.482	5	10.496	4.419	.001
Colorbw	1.691	1	1.691	.712	.999
Visaudio	39.308	1	39.308	16.549	.001
Tape	2.682	1	2.682	1.129	.288
Shot	10.535	2	5.267	2.218	.108
2-way Interactions	13.069	9	1.452	.611	.999
Colorbw visaudio	3.559	1	3.559	1.498	.219
Colorbw tape	.035	1	.035	.015	.999
Colorbw shot	3.407	2	1.703	.717	.999
Visaudio tape	.934	1	.934	.393	.999
Visaudio shot	4.619	2	2.310	.972	.999
Tape shot	.034	2	.017	.007	.999
3-way Interactions	9.702	7	1.386	.584	.999
Colorbw visaudio tape	.029	1	.029	.012	.999
Colorbw visaudio shot	6.082	2	3.041	1.280	.278
Colorbw tape shot	2.830	2	1.415	.596	.999
Visaudio tape shot	1.068	2	.534	.225	.999
4-way Interactions	47.316	2	23.658	9.960	.001
Colorbw visaudio tape shot	47.316	2	23.658	9.960	.001
Residual	1425.140	600	2.375		
Total	1547.708	623	2.484		

Multiple $\underline{r}^2$ = .034

Table 14

Cell Mean Confidence Scores, Variances and Marginal Means for the Emotional Segments

			Shot						
			Head Only		Body Only		Head & Body		
Tape			Audio and Visual	Visual Only	Audio and Visual	Visual Only	Audio and Visual	Visual Only	
Color	I	$\bar{x}$	7.22	6.21	6.14	6.14	6.83	6.02	6.43
		s^2	2.33	1.89	2.19	2.19	2.04	2.55	
	II	$\bar{x}$	6.91	5.83	6.27	5.65	6.63	5.72	6.17
		s^2	2.34	3.77	2.21	4.47	1.86	3.54	
Black and White	I	$\bar{x}$	6.79	6.58	6.98	5.69	5.70	6.23	6.33
		s^2	1.93	1.25	2.44	3.73	1.85	1.79	
	II	$\bar{x}$	6.94	5.46	6.00	5.87	6.46	5.31	6.01
		s^2	1.94	3.34	1.80	2.36	2.81	1.06	
Marginal means			6.97	6.02	6.35	5.84	6.41	5.82	
			6.50		6.10		6.12		

6.30

6.17

Transcript Only

Tape	
I	6.43
	4.24
II	5.96
	5.20

Audio Only

Tape	
I	6.33
	2.34
II	6.64
	1.11

audio & visual = 6.58
visual only = 5.89
grand mean = 6.25

*Note: . See Table 2 (page 60) for sample sizes.

Table 15

ANOVA for Emotional Confidence

Source of Variation	Squares	DF	Square	F	Probability Level
Main Effects	98.141	5	19.628	8.255	.001
Colorbw	2.068	1	2.068	.870	.999
Visaudio	69.560	1	69.560	29.254	.001
Tape	11.964	1	11.964	5.031	.024
Shot	18.199	2	9.099	3.827	.022
2-way Interactions	20.041	9	2.227	.936	.999
Colorbw visaudio	1.120	1	1.120	.471	.999
Colorbw tape	.030	1	.030	.013	.999
Colorbw shot	4.789	2	2.395	1.007	.367
Visaudio tape	6.743	1	6.743	2.836	.089
Visaudio shot	7.034	2	3.517	1.479	.227
Tape shot	.780	2	.390	.164	.999
3-way Interactions	16.933	7	2.419	1.017	.418
Colorbw visaudio tape	1.076	1	1.076	.452	.999
Colorbw visaudio shot	6.388	2	3.194	1.343	.261
Colorbw tape shot	1.225	2	.613	.258	.999
Visaudio tape shot	8.859	2	4.429	1.863	.154
4-way Interactions	23.338	2	11.669	4.908	.008
Colorbw visaudio tape shot	23.338	2	11.669	4.908	.008
Residual	1433.793	603	2.378		
Total	1592.246	626	2.544		

Multiple $\underline{r}^2$ = .062

Emotional. Examination of Table 15 indicates a significant effect for shot ($p < .025$). The marginal means in Table 14, coupled with subsequent t tests reveals the source of this effect. The head-only mean of 6.50 is significantly larger than both the body-only mean of 6.10 ($t = 2.64$, $df = 413$, $p < .01$) and the head-and-body mean of 6.12 ($t = 2.55$, $df = 425$, $p < .025$).

Visaudio

Factual. Observers who based their judgments on both audio and visual information indicated a mean confidence level of 6.20 in their accuracy judgments, while those who had visual information only had a mean confidence of 5.68. Examination of Table 13 indicates that this difference is significant ($p < .001$). Visaudio did not enter into any significant interpretable interaction with the other variables.

Emotional. Observers in the audio and visual conditions reported a mean emotional confidence level of 6.58 while those who received visual information only had a mean of 5.89. Table 15 indicates that this main effect was significant ($p < .001$). Visaudio did not enter into any significant interpretable interactions with the other variables.

Color/Black and White (colorbw)

Factual. Examination of Table 13 indicates no main effect or interpretable interactions for colorbw on factual confidence.

Emotional. Examination of Table 15 indicates no significant main effect on interpretable interactions for colorbw on emotional confidence.

Transcript-Only and Audio-Only

Factual. Observers in the transcript-only conditions had a mean confidence score of 6.07, while observers in the audio-only condition had a mean of 6.02. This difference is not significant (t = .22, df = 84).

The visual-only mean confidence score of 5.68 for all 12 visual conditions did not differ significantly from the audio-only mean of 6.02 (t = 1.31, df = 350). Within the visual-only conditions, just the body-only mean of 5.38 was significantly less than the audio-only mean (t = 2.31, df = 134, $p < .05$).

The visual and audio mean confidence score across all 12 conditions was 6.20, which does not differ significantly from the audio-only mean of 6.02 (t = .72, df = 358). Within the visual and audio conditions, none of the shot means differed significantly from the audio-only mean.

Emotional. Observers in the transcript-only conditions had a mean confidence score of 6.20 while observers in the audio-only conditions had a mean of 6.49. This difference is not significant (t = .75, df = 85).

The visual-only mean confidence score of 5.89 for all 12 cells is significantly less than the audio-only mean

of 6.49 (t = 2.36, df = 350, p < .02). Comparing the audio-only condition mean of 6.49 with the visual-only conditions, shows both the body-only mean of 5.84 (t = 2.35, df = 134, p < .025) and the head-and-body mean of 5.82 (t = 2.47, df = 147, p < .02) to reflect less observer confidence.

The visual and audio mean confidence score across all 12 conditions was of 6.58 which does not differ significantly from the audio-only mean of 6.49 (t = .35, df = 358), nor does the audio-only mean differ from the three means from the visual and audio conditions of the shot variable. The head-only condition had a mean of 6.97 (t = 1.75, df = 141); the body-only had a mean of 6.35 (t = .52, df = 147); while the head-and-body visual and audio mean was 6.41 (t = .29, df = 146).

The correlation between accuracy and confidence for the factual segments was .063, while for the emotional segments it was .061. Due to the large samples on which these overall correlations are based, both are significant (p < .05). The correlations between accuracy and confidence for each cell in the design for both the factual and emotional segments appear in Appendix H.

Footnotes

[1]The analysis of variance assumes that error variances within treatment conditions are equal. Hays (1963, p. 379) points out that this assumption can be violated without serious consequences if the sample size in each condition is equal. An examination of Table 2 reveals that, while the sample sizes in each cell are close to being equal, they are not exactly so. The resulting problem is potentially serious since the basic unit of analysis for both the factual and emotional accuracy are proportions. Winer (1971, pp. 399-400) points out that in this situation, the individual cell means and variances are not independent. As the particular cell mean deviates from .50, the cell's variance tends to increase. Thus only in the situation in which the overall null hypothesis were true (there were no treatment effects and all cell means were equal) would the homogeneity assumption likely be met. Winer (1971) recommends stabilizing the variances by transforming the raw proportions into a value two times the arcsine of the square root of their original value. Analyses of variance on the accuracy scores were performed both on the raw proportions and on these scores after the recommended transformation had been performed. Both of these analyses yielded almost identical results. Since using the transformed scores results in interpretation difficulties, the results to be presented here are from the analysis of variance performed on the raw proportions. The same analyses for the transformed scores appear in Appendix G.

[2]Since the dependent variables were obtained from the same observers and are thus correlated in the same unknown way, the significance tests reported are not independent of one another and no exact probability that at least one of them will exceed a particular critical level on the null hypothesis can be calculated (see Kerlinger and Pedhazur, 1973, p. 352). An alternative procedure to the one used here would be to have conducted a multivariate analysis of variance (MANOVA) which would have allowed comparison on the dependent variables taken as a set. This was not done because while the dependent variables are not statistically independent of one another, they are conceptually distinct. For example, a major purpose of this experiment was to examine emotional versus factual accuracy across levels of the various independent variables and MANOVA would not directly address such comparisons.

[3]The analyses of variance were performed using subprogram "ANOVA" from version 6.0 of the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, and Bent, 1975). When a factorial analysis of variance is

performed on data with unequal cell sizes, as in the case of this experiment, the effects of the independent variables are not orthogonal (i.e., joint effects will exist). The technically correct solution to this problem would be to perform a series of Anovas, varying the order in which the independent variables were entered. A series of Anova source tables would result, each of which would be slightly different depending on which variables were entered first and consequently which variable was assigned the joint effects. The SPSS Anova program performs an unweighted analysis of the cell means which results in just one source table. This method is an approximation technique in which each cell is treated as if it contains only one observation. An estimate of the extent to which this approximation technique deviates from what would have been attained if cell sizes had truly been equal is obtainable by comparing the true total sums of squares with the summed sums of squares from each source of variance in the analysis used in the approximation technique. (The true sums of squares is found by multiplying the variance of the total sample by sample size.) The true SS_T for factual accurary is 22.88, while the SS_T in the analysis performed is 23.43. For emotional accuracy the figures are 15.37 and 15.02. For factual confidence and emotional confidence, respectively, the figures are: 1575.17 and 1547.71; 1608.84 and 1592.25.

These comparsions indicate that the source tables to be presented here are nearly identical with what would have been reported if the sample sizes had been equal for every cell.

[4]Inasmuch as this experiment is exploratory, it is desirable to make some comparisons which were not specifically planned. Scheffe (1953) has suggested a test that is appropriate for making any and all comparisons that may be suggested by the values of the cell means themselves with a probability of 1 minus alpha that all statements concerning the significance of these comparsions will be true. This procedure involves performing a standard t test (using the MS_W from the overall Anova) but requiring a critical t equal to or greater than the square root of the number of cells in the experiment minus one, times the critical t required for significance from the original Anova. The critical F for an alpha of .05 in this experiment (df = 27,600) is about 1.6. This yields a critical t for Scheffe's test of about 6.4. For an alpha of .01 the t required is about 7.0. Scheffe's test, then, would require an extremely large difference in cell means to yield a significant result since there are so many possible comparisons against which to protect an overall experiment-wise alpha. Using such a conservative test would probably result in failing to detect many "true" relationships which exist in the population. Consequently, the procedure which will be followed here will be to perform the t

tests and report the resulting t values and the two-tailed probability. Any comparison which exceeds the .05 level two-tailed will be considered significant. The reader should be alerted to the fact that this procedure does not provide experiment-wise error protection and it makes the probability of committing Type I error on any given test considerably higher than if Scheffe's procedure were strictly followed.

[5]For both factual and emotional confidence, four-way interactions were found. These interactions were extremely difficult to describe verbally and consequently such a description will not be attempted. The nature of these interactions are represented numerically by the 24 cell means in Table 12 and 14. These interactions will be discussed in the next chapter.

Chapter IV

DISCUSSION

The purpose of this chapter is to interpret the results of the experiment, to discuss the practical and theoretical implications, to evaluate the procedures, and to suggest future research.

Accuracy

Shot

The rationale for this exploratory experiment draws heavily on the work of Ekman and Friesen (1969a; 1974). A major purpose was to replicate their (1974) research using samples of lying and truthing behavior from another kind of dishonest interaction in addition to replicating under conditions approximating their original procedures. For emotional accuracy, the body-only condition resulted in significantly higher accuracy scores than did the head-only condition. Examination of the marginal and cell means in Table 4 indicates that while this finding exists as a main effect, it is the result of the accuracy scores in the visual-only conditions of tape I. It is perplexing that this effect is only found for tape I. A reasonable possibility is that only some subjects emitted more body than

head leakage when they were exposed to the stressful slides. By chance these individuals may have ended up lying on tape I and truthing on tape II, and consequently, accuracy would be higher on tape I.

For emotional accuracy, then, the results here provide relatively unambiguous support for the Ekman and Friesen (1974) head-body finding.[1] For factual accuracy a different situation exists. The head-only condition observers were significantly more accurate than were the body-only observers. Within the visual-only conditions, this difference approached but failed to reach significance ($t = 1.58$, $df = 202$, $p < .10$). This finding makes questionable the generalizability of the Ekman and Friesen (1974) head-body finding. The body may be a better source of cues which are indicative of deception than the head only when the deception involves the concealment of emotional responses. When the lying and truthing behavior involve statements about factual events, greater accuracy resulted from judgments which were based on the head only. It is thus likely that the head is the best source of information in these situations. Probably few body cues were emitted by subjects who were lying or telling the truth about the factual events.

For emotional accuracy, if the body provides the greatest source of cues for detecting deception concerning reports of emotional feelings (as the results of both the Ekman and Friesen [1974] study and this experiment suggest), then the presence of the head could distract observers from

noticing these cues. It is natural to watch the head of someone who is talking, rather than concentrating exclusively on the body as observers in the body-only condition were forced to do. Observers in the head-and-body condition, with a mean of .497, were less accurate than those in the body-only condition, who had a mean of .520. This difference approaches but fails to reach significance both as a main effect (t = 1.56, df = 419, p < .15) and within the visual-only conditions, where the head-and-body mean was .491 and the body-only mean was .528 (t = 1.74, df = 206, p < .10).

For factual accuracy, the body apparently did not provide the best source of information about veracity, since observers were more accurate in the head-only condition. The accuracy level in the head-and-body condition (.545) was almost identical to the head-only (.537) as a main effect, and within the visual-only conditions these means were identical (.480). If observers who were shown the head-and-body were distracted from watching the body, this distraction would not detrimentally affect their accuracy score unless the body were a good source of information about truthfulness. In fact, it would be advantageous for the observer to be distracted from paying attention to an area of the body which was not emitting cues indicative of deception. Observers in the head-and-body condition could observe less head detail than their counterparts in the head-only condition, yet their accuracy scores were just as high. Several possibilities could account for this finding. Accurate judgments could be

based upon gross head movements, rather than head detail, and if so, both the head-only and the head-and-body observers may have been able to notice these movements. Another possibility is that the body may have provided enough information about deception to counteract the information lost by head-and-body observers because of the unavailability of head detail. A final possibility is that inconsistencies between the nonverbal behaviors being emitted by the head and the body, which would have only been available to the head-and-body observers, may have enhanced their accuracy scores. The design of this experiment does not provide evidence bearing on these explanations.

It is clear that emotional accuracy was higher when judgments were based on the body-only than when based on the head-only, whereas for factual accuracy an opposite result was found. The head-and-body accuracy scores are consistent with the interpretation that for emotional accuracy the best source of cues is the body, but that for factual accuracy the body _per se_ is not the best source of information. The overall pattern of results on the shot variable can be taken as a potentially important refinement and extension of the Ekman and Friesen head-body hypothesis about leakage and clues to deception.[2]

Before turning to a discussion of the other variables in the study, two qualifications about the meaning of these results need emphasis. First, there is a temptation to conclude that the head-body results presented here provide

strong evidence that the body is the best source of information about emotional deception, while the head is the best source of information about factual deception. Although there is some basis for this conclusion, another possibility cannot be completely discounted. As was noted earlier, cues indicative of deception are useful to observers only if they are both noticed and correctly attributed. Observers' beliefs and stereotypes about how liars and truthers behave may be grossly inaccurate. Thus, for factual accuracy, the body-only observers may have been less accurate than their head-only counterparts because some bodily cues actually indicative of deception were incorrectly attributed to truthing, and/or vice versa. Thus, the results of this experiment do not necessarily mean that the "best" source of cues denoting deception are those areas where accuracy is highest. If observers believed that many body movements indicated lying and few body movements indicated truthing, and the opposite were true, accuracy scores for body-only observers would be low even though cues indicative of deception were plentiful. Conversely, fewer cues might exist in the head-only condition, but these cues might be more consistent with observers' beliefs about the behavior of lying individuals, and consequently, relatively higher accuracy scores would be likely.

A second qualification concerns the familiarity sample of subjects' behavior which was presented to observers. In this segment, observers knew that subjects were always telling the truth. Thus, it is not clear if observers saw

more cues indicative of deception in one condition than in another, or if the deception behavior deviated from the truthful familiarity sample more in some conditions than in others. In other words, observers may not have known if specific cues were indicative of deception or not, but if a particular behavior were different in the test segment than in the familiarity sample, they may have correctly inferred that this difference was a result of lying.

Several factors suggest that this latter possibility is at least partially accounting for the results. If observers were able to identify cues indicative of veracity more accurately from some parts of the body than from others, Ekman and Friesen (1974) should have been able to obtain the predicted head-body differences even when no familiarity sample were provided. As was noted in Chapter I, they could not. Furthermore, the inconsistencies and contradictory findings in previous studies employing content analyses of lying and truthing behavior add support to this interpretation. Perhaps lying behavior is highly idiosyncratic. If so, the best way to detect lying would probably be to compare samples of known truthful behavior with samples of doubtful veracity. Behavior which deviated from known truthing behavior would then result in an inference of lying. To the extent that this reasoning is correct, the results of this study do not provide evidence about which categories of behavior contain specific cues indicative of veracity, but rather which

categories contain behavior which deviates the most from truthing to lying.

Colorbw

No accuracy differences were observed for the colorbw variable. Several features of the overall pattern of results, however, dictate caution in concluding that observers were equally accurate when watching the tapes in color and in black and white. The colorbw variable would be most likely to make a difference in the head-only condition where observers were viewing a close-up of the subject's head. Within this condition, all eight pairs of cell means (four pairs for each dependent variable) were higher in color. Within the other levels of the shot variable there is no consistent pattern of either color or black-and-white resulting in different levels of accuracy. For factual accuracy, comparing the head-only color mean of .550 with the black and white mean of .523 yields a non-significant t to 1.12 (df = 208). A similar finding results from the same comparison for emotional accuracy, where the color mean was .493 and the black-and-white mean was .457 (t = 1.73, df = 206, p < .10). Although not technically appropriate, combining the two accuracy dependent variables results in a color head-only mean of .521 and a black and white mean of .490. This comparison is significant (t = 1.98, df = 416, p < .05). This finding, plus the consistency of the pattern of color resulting in greater accuracy than black and white in the head-only

conditions, suggests that for these conditions color may improve observers' ability to detect deception.

Whether there are color cues such as facial flush which result in greater accuracy in the color conditions or whether observer interest or motivation is higher in color, which in turn results in greater accuracy, is unknown. Data bearing on this latter possibility exist in the form of observer responses to a questionnaire item which asked the extent to which they found the task of attempting to identify the veracity of the subjects interesting. Responses to this question indicate no relationship between format and observers' reports of their interest in the task. For the head-only conditions, black-and-white observers indicated a mean interest of 6.19 while color observers had a mean of 6.16 (t = .07, df = 210).

Part of the rationale for including the colorbw variable was to increase the potential applicability of the findings, specifically to the court system. If, in the future, trials or portions of trials are presented to jurors on television, the results of this experiment provide some evidence that jurors may be able to detect lying more accurately in a color than in a monochromatic format.

Paralinguistic Cues

Perhaps the most perplexing findings in this experiment are the results associated with the paralinguistic cues. For both factual and emotional accuracy, observers who were

exposed to subjects' behavior in an audio-only format did not differ significantly from those in the transcript-only conditions, although for emotional accuracy, transcript-only observers approached being significantly more accurate than those in the audio-only conditions ($t = 1.71$, $df = 85$, $p < .10$). Several possibilities could account for this lack of differences. Perhaps there were no paralinguistic correlates of lying which permitted accurate inferences by observers. A second possibility is that the paralinguistic characteristics associated with the audio format distracted observers from noticing aspects of the verbal content which enhanced the accuracy of those in the transcript conditions. A final possibility is that the paralinguistic characteristics of the behavior confused observers. For example, in some cases they may have inferred erroneously that nonfluencies were a result of lying when, in fact they were merely a result of the subjects' general nervousness as they were being interviewed by the police officer. The situation in which the stimulus tapes were made was designed to create high saliency and ego-involvement on the part of subjects. The effort to heighten saliency and involvement may have been too successful; that is, many subjects were visibly nervous during their interviews both when they were lying and when they were truthing. To some extent observers in the audio-only condition may have correctly inferred veracity based on paralinguistic information, but errors resulting from subjects' general anxiety level may have reduced their accuracy scores.

The strongest finding in the entire experiment was that, for factual accuracy, observers in the visual and audio conditions were more accurate than those in the visual-only conditions. Several factors mitigate against concluding that the paralinguistic cues are providing observers with the best information from which to infer veracity. First, despite efforts to reduce the variability of the content of the subjects' lying responses, there was still more variability than when they were truthing. Also, subjects in some cases provided more detailed responses when truthing than when lying. The fact that the factual accuracy for the transcript-only observers were so high (.625) suggests that the verbal content itself did contain information from which to judge veracity accurately. This content, of course, was not available to visual-only observers. Furthermore, an error was made in creating the factual accuracy questions which the police officer asked the subjects. One question asked if Mr. Bostick, the person they had seen on the videotape before their interview, had the right to appeal his conviction. The last question in the factual segment asked what Mr. Bostick's reaction was to being told that all he had to do to appeal his conviction was to sign a form. Thus, this question suggested the answer to the earlier question. No subject appeared on either tape answering both of these questions. However, it is possible that some observers noticed this inconsistency between the questions and correctly inferred

that any subject who said that Mr. Bostick did not have the right to appeal his conviction was lying.

No such problem with either the questions or the variability of the answers exists for emotional accuracy and no main effect or interactions were found for visaudio. Whatever information the audio band contains which was relevant to the identification of veracity apparently was not useful to observers. The audio band may have distracted observers from noticing visual cues, or vice versa. Or whatever aid the paralinguistic information provided observers may have been cancelled by inaccurate judgments resulting from misinterpreting these cues. In any event, visual-only observers were just as accurate as visual and audio observers.

Comparing the audio-only conditions with the visual-only conditions results in a non-significant trend in favor of greater emotional accuracy for the audio-only observers. The highest visual-only accuracy within the shot variable was found in the body-only condition. This mean of .528 was significantly higher than the audio-only means of .473 (t = 2.09, df = 135, $p < .05$). It is unknown if there were more cues indicative of deception emitted by the body than by the voice, or if perhaps observers' beliefs about which cues were indicative of lying were more accurate for the body than for the voice.

Emotional Versus
Factual Accuracy

Summing across all 28 cells in the experiment, factual accuracy was considerably higher than emotional accuracy (.541 vs. .499). It should not be concluded from this comparison, however, that more cues indicative of deception existed for the factual segments or that these cues were more consistent with observers' beliefs about which cues are associated with deception than cues which were emitted by subjects during the emotional segments. These factors mentioned which contributed to the audio and visual observers having higher factual accuracy scores than visual-only observers are also likely to be accounting for the overall factual accuracy scores being higher than the emotional accuracy scores. The visual-only factual accuracy mean of .467 was actually lower than the emotional accuracy mean of .491. It seems clear that the verbal content of the audio and visual conditions is also accounting for factual accuracy being higher than emotional accuracy.

A further problem which makes comparisons between the factual and emotional accuracy scores difficult is a possible order effect. The factual segments were always presented first. Consequently, subjects may have become bored with the task by the time the emotional segments appeared, and thus have been more accurate for the factual segments or alternatively, they could have become better at making the judgments with practice and thus could have been more accurate for the

emotional segments. Randomizing the order of presentation would probably have been desirable but this would only have been possible if observers could have been run in smaller intact groups or individually. It should be noted that since the order of presentation was a constant, it would not influence comparisons between the experimental conditions unless order interacted with the manipulated variables.

Confidence

For both factual and emotional confidence there was a highly significant main effect for visaudio ($p < .001$ for each). Observers who both heard and viewed subjects' behavior were more confident in the accuracy of their judgments than their counterparts in the visual-only conditions. The audio-only observers were significantly more confident for the emotional segments than either the body-only or the head-and-body observers in the visual-only conditions. A similar pattern of results was found for the factual segments except only the body-only mean within the visual-only cells was significantly less than the audio-only mean. Apparently the presence of the audio band increases observers' degree of confidence in their judgments and observers are more confident when they have just audio information than when they have just visual information, except when the visual information is in the form of a close-up of the head for the emotional segments, and when the head is available in any form for the factual segments.

For emotional confidence, there is a significant ($p < .025$) main effect for shot with the head-only conditions having a significantly higher mean than either the head-and-body or the body-only. No similar main effect exists for factual confidence; however, a t test indicates that observers in the head-only condition were significantly more confident than were those in the body-only condition ($t = 2.38$, $df = 413$, $p < .02$). Thus the overall pattern of results for shot on both confidence dependent variables indicate that observers are most confident when judgments are based on a close-up of the head, next most confident with the head and body, and least confident when judging the veracity of a headless body.

A color vs. a black-and-white format apparently has no effect on observer confidence. Even combining the two dependent variables and examining the head-only level of the shot variable fails to yield a significant color vs. black-and-white comparison ($t = 1.16$). It should be noted that there is a slight overall trend for greater confidence resulting from the color conditions.

A particularly difficult finding to interpret is the fact that significant four-way interactions were found for both factual and emotional confidence. The procedure for marking the scales on which observers indicated their degree of confidence in their accuracy judgments was not explained on the questionnaire in detail. Consequently in several conditions, observers asked questions about how to mark these

measures. Deviations in the nature of these questions and in the manner in which they were answered by the experimenters may have created systematic biases towards indicating more or less confidence in some cells than in others. Inasmuch as it is difficult to imagine theoretical reasons for such complex interactions, it seems plausible that some random factors were contributing to confidence ratings. The lack of specificity in the directions for filling out the confidence scales may have contributed to this. More detailed instructions probably should have been provided on the questionnaire.

The relationship between accuracy and confidence was a very weak one. Within most of the cells of the design for both the emotional and factual segments, the correlations do not exceed what would be expected by chance (see Appendix H). Overall, including all 730 observers, there were modest correlations of .063 for the emotional segments and .061 for the factual segments. Because of the large sample size, these correlations are significant. These small correlations are consistent with the results of the only other study which has examined the relationship between accuracy and confidence (Matarazzo et al., 1970). This study failed to find any relationship between these variables.

Observers probably have beliefs about which categories of behavior provide the best information about deception. The weak association between accuracy and confidence may be evidence that those beliefs are largely wrong. When observers

thought they were being provided with a category of behavior from which cues to deception were likely to be emitted, they were more confident of their judgments than when other categories of behavior were presented. For example, the eyes have been thought to provide accurate information about deception. When observers were provided with a close-up view of the eyes, as in the head-only conditions, they indicated more confidence than when the eyes were either not seen at all, as in the body-only conditions, or were difficult to see, as in the head and body conditions. Thus eye observability may have at least partially accounted for the confidence ratings varying across levels of the shot variable. If eye behavior were unrelated to veracity, this would contribute to low correlations between accuracy and confidence.

Problems with the Study

The purpose of this section is to describe weaknesses in this experiment and to suggest ways in which they might have been corrected. This presentation will be divided into two sections: weakness associated with creating the stimulus tapes; and weaknesses associated with the data collection procedures.

Creating the Tapes

The decision to use criminal justice majors as subjects stemmed from the need to create the perception that important consequences would result from detection and because of the need to replicate the Ekman and Friesen (1974)

procedures as closely as possible. Criminal justice is concerned with lie detection, and it is possible that many of the subjects who participated had more knowledge about lie detection than the average person. This knowledge could have affected the manner in which these particular subjects approached the task of attempting to deceive the officer. Using this subject population obviously limits the confidence with which the findings reported here may be generalized to other lying and truthing individuals. There is no way of knowing, at this point, the extent to which these subjects may have had knowledge about lying and its detection or the extent to which this knowledge did affect their own behavior during their interviews. It probably would have been desirable to have attempted to systematically gather information from subjects about these possibilities.

The attempt to create the perception in subjects that serious negative consequences for them would result if they were detected was apparently quite successful. In general, subjects were quite anxious about their performance on this task. Unfortunately, this anxiety may have been observably manifested both when subjects were lying and truthing. Thus the overall accuracy scores may have been reduced by the erroneous inference by observers that these indications of nervousness were a result of lying when they were actually indicative of nervousness resulting from the situation itself. The two color cameras, the use of a television studio, the police detective and other characteristics of the taping

environment probably contributed to this. Furthermore, the familiarity segments which were shown of each subject before observers saw the "test" segments, were the first segments taped. Thus they were taped when any general anxiety subjects were experiencing was likely to be at a maximum. It probably would have been better to have taped the truthful "baseline" segments at a latter point in the interview to increase the likelihood that subjects would appear relaxed during this segment.

It has been implied here that the attempt to create the belief in subjects that serious consequences would result for them if they were detected was highly successful. The only evidence for this, however, is the subjective impressions of the researchers. No subjects indicated suspicion about the cover story at any point and all appeared to take their task very seriously. This subjective impression should have been verified with a more systematic assessment of subjects' beliefs about their participation. An item on the subjects' post-interview questionnaire should have asked them about their beliefs about the consequences resulting for them from detection.

The police detective who interviewed the subjects asked the questions in a non-threatening way. It may have been desirable for him to take a more adversarial role; that is, to have verbally aggressed against the subjects to try and upset them. Lying subjects may have become more upset

than truthful subjects and more cues to deception may have existed, thus providing a better test of the hypotheses.

Data Collection

One major problem associated with data collection was the use of intact groups. This procedure created two related problems; statistical, which are relatively minor; and rival hypotheses for the results, which are serious.

The most important assumption of all inferential statistical techniques is that the sample is random and that the observations are assigned to particular experimental conditions independently. Using intact groups violates this assumption with the result that the degrees of freedom in the analyses of variance are inflated. Instead of having a total of 623 total degrees of freedom, there would be 23. Each cell would necessarily be treated as if it had just one observation. Consequently, no within group variance could exist and the error term with which to test the main effect hypotheses would be the mean square associated with the interaction term. This procedure requires the assumption that there will be no interaction and thus the interaction MS is assumed to be error. The F's required for significance would be considerably larger for all tests. Overall, this procedure, while statistically accurate, would be conservative and especially inappropriate given the exploratory nature of this experiment.

Of a more serious nature is the threat to the internal validity of the experiment that the use of intact groups

creates. Each intact group had a short history (about 2-3 weeks) together as members of an introductory communication class. In these classes, communication research is sometimes discussed. These discussions could affect the participants' attitudes toward research which, in turn, could have influenced their motivation and interest in the task. It is further possible that class discussions of nonverbal communication could have influenced the observers' abilities to detect deceptive communication. If observers had been assigned at random, these possibilities would have been much less serious because it would be likely that observers with a common history would be represented about equally in each of the cells of the design.

A further problem was that the treatments themselves were, for the most part, administered to all observers within a cell at the same time. The existence of "intra-audience" effects has been well documented (cf., Hocking, Margreiter, & Hylon, in press). Individuals' evaluations of messages and sources are influenced by the evaluations and observable responses of those around them when the message is received. Observers in this experiment were explicitly told to make their own judgments of subjects' veracity and not to talk at any point while the tape was being played. In spite of these precautions, some observers occasionally made comments which could have affected the responses of other observers. In at least one group it was necessary for E_1 to ask for such comments to cease.

Each of these factors is a rival hypothesis for the results of this experiment. So, for example, it is possible that the group which had the highest factual accuracy score of .682 (head-and-body, audio and visual, color, tape I) had attitudes and knowledge which were unique to them as a group which increased their accuracy relative to other groups. Or the group which had the lowest emotional accuracy score of .417 (head-only, visual-only, black and white, tape I) may have had (a) particularly influential class member(s) emitting nonverbal, and perhaps even verbal, responses which caused other observers to error in judging veracity. This problem would have been solved if observers were assigned independently at random to the 28 cells of the design and exposed to the stimulus in isolation (i.e., one at a time).

It should be noted that if the use of intact groups were seriously influencing the results of this experiment, it would be unlikely that the main effects which have been reported would be as systematic as they are. It is fortunately improbable that the use of intact groups had a major affect on accuracy judgments. The point, however, is not that this procedure did or did not have an affect on the results, but rather that the affect is unknown. If a completely random observer assignment procedure had been used and if observers had been run in isolation, such an unknown would not exist.

Observers were run over an eight-day period. Thus it is likely that many of the observers who had already

participated interacted with others who had not yet participated. If these interactions included discussions of this research, it is possible that later participants' knowledge of the procedures influenced their performance. Since the 28 cells were assigned (almost) at random to the particular classes and days, it is not likely that this factor could have influenced the results systematically.

Each of these problems represents clear weaknesses in this research. Confidence in the results needs to be tempered accordingly.

Future Research

A reoccurring theme throughout this paper has been the relationship between detection accuracy and the amount of information contained within a particular presentational format. It was pointed out that more information within a level of an independent variable could either increase accuracy by providing additional cues from which observers could correctly infer veracity, or decrease accuracy by distracting observers from noticing cues to deception. Each level of each variable in this experiment might be classified according to its total amount of information. For example, the visual-and-audio conditions contain more information than do the visual-only conditions; the head-and-body conditions contain more information than the body-only conditions; and the color conditions contain more information than the black-and-white. If these levels could be quantified along a single dimension of "total

information" for each subject, the relationship between accuracy and information could be more precisely examined. It might be found, for example, that the relationship is curvilinear: additional information increasing accuracy to a point, but decreasing it beyond this point.

Probably no variable in the Communication discipline has received as much research attention as "credibility." The present study also has relevance for this variable. Situations in which a dichotomous judgment of veracity are made by observers are probably rare. Situations in which the relative credibility of a source is judged are quite common. These behaviors from which observers in the present study infered lying, would probably also lower observers' perceptions of credibility in situations where a dichotomous judgment of veracity was unlikely. An interesting research approach might be to present the stimulus tapes from the present study to observers and ask them to assess the credibility of each, then relate these evaluations to judgments of veracity made by comparable observers.

In detecting deception in everyday life, the extent to which observers rely on observable behaviors is unknown. Before an inference of lying is made, certain situational variables probably need to create the suspicion of deception. The most obvious situational factor would be the observers' perceptions of gains to be made by the source by presenting particular verbal statements. A criminal suspect obviously has something to gain by denying the crime, thus the police

officer is suspicious of the denial. The subordinate has gains to be realized by flattering the superior, thus the superior is suspicious of the flattery. The same subject behaviors emitted in different situations would be likely to result in different observer judgments. For example, speaking non-fluencies in one situation might result in an inference of lying, in another of nervousness, and in another of low credibility. The cause to which the behaviors were attributed by observers would probably affect the type of evaluation they would make. Attribution theory may provide a useful perspective from which to address the role of situation in detecting deception. This approach would involve manipulating situational characteristics to affect the cause to which subject behaviors were attributed, and examining the effect on judgments of veracity.

Nonverbal communication has received extensive research attention in recent years. The importance of the nonverbal component of communication messages has been emphasized by numerous writers to the point that such proclamations have become accepted truisms. Mehrabian (1968) has gone so far as to say that 93% of the impact of messages comes from the nonverbal dimension. The results of the present experiment suggest that for accurately detecting deceptive communication, the nonverbal component is less important than the verbal content. For both the factual and emotional segments, accuracy was higher in the transcript-only conditions which contained only the verbal content, than

in conditions which contained either visual or paralinguistic information. These results are consistent with the findings of Maier and Thurber (1968) that readers were more accurate than were either listeners or watchers in detecting deception. If, as these results suggest, there are circumstances in which nonverbal behaviors can distract observers from accurately decoding the verbal meaning in a message, there may be an important implication for the court system. Proponents of the use of videotape have suggested that videotaped records of trial proceedings would be superior to the traditional transcript. One advantage, they say, would be that during deliberations jurors could review testimony on television, thus being able to examine the nonverbal behavior of the witness in addition to having the verbal content. These results suggest that the nonverbal dimension of the behavior could distract the jurors from carefully scrutinizing the verbal material. Research examining this possibility in a legal context would be of value.

Ekman and Friesen (1974) make the obvious suggestion that their results need to be replicated with additional subjects and observers. This is true particularly with respect to subjects, since both Ekman and Friesen (1974) and the present experiment used subjects from very homogeneous populations. Replications using stimuli from other kinds of dishonest interactions would also be of value.

Individual idiosyncracies may be contributing to the confusion associated with the results of those studies

employing content analyses. A line of research suggested by the results of this experiment would involve a modification of the content analytic approach. If, as has been suggested above, there are few specific lying behaviors which are general to many deceivers, it may be that categories of behavioral deviations from lying to truthing are more general. For example, Knapp et al. (1974) found no differences between the number of "eye contact units" between liars and truthers. Perhaps some lying individuals have more of those units than when they are truthing, and others less. The category itself still could be a good source of information if deviations within the category for each individual were examined, rather than summing the number of behaviors across individuals. Such a lying to truthing deviation content analytic study on the stimulus tapes from the present experiment is intended. This is not to suggest that content analyses looking for specific behaviors across individuals should be abandoned. Possibly when these analyses are conducted using stimulus materials created in situations which have non-trivial consequences for the subjects, more consistency in the findings will result.

As discussed earlier, it is impossible to infer from the results of this experiment which categories of observable behavior contain the best cues to deception since observers' beliefs about the behavioral correlates of veracity are influencing their accuracy. I have been able to find no study in the deception literature which examines beliefs about how

liars behave. Maier and Janzen (1967) asked observers to list reasons for their judgments of lying or truthing and found no relationship between accuracy and the reasons given. As the authors point out, observers may not have known why they were making particular judgments and generated reasons in a post-hoc fashion to justify their decisions. Research examining beliefs about what observable behaviors are associated with lying would probably be of value. The materials for one such study exist within the present experiment and I intend to pursue this question. By combining information about the total number of judgments of true and false for each subject in each condition, with a content analysis of each subject's behavior, it should be possible to identify the specific behaviors from which observers are inferring veracity.

The data generated by this experiment will also allow several questions related to other aspects of deceptive communication to be addressed. Demographic information was collected from all observers. It is thus possible to examine both the degree of detection accuracy and confidence in judgments by various subgroups within the overall sample. It will be possible to identify the characteristics of these observers who were most adept at judging veracity.

Summary

The following is a summary of the major results.

1. The Ekman and Friesen (1974) finding that observers who based judgments on the body-only were more accurate

than those who based their judgments on the head-only was replicated for the emotional segments. For the factual segments, a strong trend in the opposite direction was observed. These results suggest that the original finding partially resulted from subjects' emotional responses to the stimulus films themselves.

2. The accuracy scores for the head-and-body conditions suggest that the head distracts observers from closely viewing the body. For the emotional segments, where the body apparently provided the best information for making inferences about deception, the presence of the heads thus reduced accuracy. For the factual segments, where the head apparently provided the best information about veracity, the presence of the head tended to increase accuracy.

3. Some evidence indicates that within the head-only conditions of the shot variable, accuracy is higher in color than in black and white. The colorbw variable had no effect on accuracy for the other conditions of the shot variable. Although it is not clear whether this effect within the head-only conditions is a result of color cues per se or a result of greater observer interest in observing the color format, it appears that observers were equally interested in viewing both formats.

4. For the factual segments, audio information resulted in greater accuracy than visual information. The fact that the transcript-only factual accuracy scores were

also high suggests that it was the verbal content of the behavior, rather than paralinguistic information, which resulted in this increased accuracy.

5. For the emotional segments, there was no relationship between visaudio and accuracy. Overall observers in the visual-only conditions were not more accurate than audio-only observers. However, within the body-only visual-only cells, the mean accuracy score was significantly larger than the audio-only mean.

6. For both the factual and emotional segments, observers who received audio and visual information were significantly more confident in their accuracy judgments than those who received visual information only. Head-only observers were more confident than head and body observers, while body-only observers were the least confident.

Footnotes

[1]Ekman and Friesen originally hypothesized that observers in the body-only condition would be more accurate than those in the head-only condition only when they were observing lying samples of the subject's behavior. Their results, however, suggest that the head-body finding might be more general than they predicted, since there was a marked, though not significant, trend in the same direction when truthful behavior was being judged. A distinction between accuracy for lying and truthing subjects has not been made here for this reason. The results of this experiment also suggest that, for emotional accuracy, the body resulted in greater observer accuracy than did the head both when subjects were lying and truthing. For factual accuracy, the head and body were about equal when truthing behavior was being judged, and the head resulted in more accuracy than did the body when lying behavior was being judged. To compare the Ekman and Friesen (1974) results with those obtained here, it is necessary to examine the accuracy scores separately for lying and truthing samples of subject's behavior. The following table does this, allowing a direct comparison in those cells which directly replicate Ekman and Friesen (1974) for both honest and deceptive segments of the subjects' behavior. That is, the means in the present experiment from the visual-only, head-only and body-only cells, collapsing on tape and colorbw, are shown.

Table 16

Comparison of the Accuracy Means and Sample Sizes* for Honest and Deceptive Samples of the Ekman and Friesen (1974) Experiment with Those From the Present Experiment

		Head			Body		
		Ekman	Hocking		Ekman	Hocking	
		(Emotional)	(Emotional)	(Factual)	(Emotional)	(Emotional)	(Factual)
Honest	$\bar{x}$	.431	.416	.387	.501	.457	.400
	n	48	107	107	55	96	96
Deceptive	$\bar{x}$	.477	.497	.573	.635	.602	.483
	n	48	107	107	55	96	96

*All observers judged both honest and deceptive segments. Thus, in this table the same observers, within each experiment, made all head judgments, while a second group made all body judgments. Each observer made four judgments in each cell of this table.

[2]It should again be noted that Ekman and Friesen (1974) used only female subjects. The present experiment used three female and 13 male subjects. Although several experiments have reported that females display more non-verbal facial expressions in non-deception contexts than do males (Buck, Miller, & Caul, 1974; Buck, Savin, Caul, and Miller, 1972), the fact that Ekman and Friesen's (1974) head-body finding was replicated for emotional accuracy suggests that the sex of the subjects did not account for the failure to replicate this finding for factual accuracy.

APPENDICES

APPENDIX A

Subject Participation Invitation Letter

MICHIGAN STATE UNIVERSITY

COLLEGE OF SOCIAL SCIENCE • SCHOOL OF CRIMINAL JUSTICE

EAST LANSING • MICHIGAN • 48824

November 12, 1974

Dear Criminal Justice Student:

The School of Criminal Justice in cooperation with the Department of Communication is conducting a research project which will attempt to identify certain personal characteristics of individuals which may contribute to their successful performance as police officers. This project compliments such research as the New York Police Training and Performance Study, the RAND Study of Police Background Characteristics and Performance, and the Systems Training Analysis of Requirements (STAR) Study. The results of this study will influence decisions regarding police personnel development programming.

We are particularly desirous of obtaining your participation in the study because of your status in the School of Criminal Justice. Some 30 students have been selected from those currently enrolled in CJ 335 and CJ 499 because of their expression of a particular interest in pursuing a career in the police profession. It is important to the research design to solicit participation from those currently enrolled in a course addressing police issues and problems and who additionally intend to pursue a career in the field.

Your participation in the study will involve just one hour of your time selected to be convenient for you. Within the next few days you will be contacted by a representative of the Department of Communication who will provide additional information.

We appreciate your cooperation in this important research effort and look forward to your participation.

Yours sincerely,

A. F. Brandstatter

A. F. Brandstatter
Director

AFB:cka

APPENDIX B

Interview Format Review Sheet

APPENDIX B

Subject number _____ Name _____

	Supposed to be	Actually was
Q 1. What was the crime which Mr. Bostick was convicted of?	_____	_____
Q 2. Did Mr. Bostick say he was sorry for his crime?	_____	_____
Q 3. Did the judge say that Mr. Bostick had the right to appeal his conviction?	_____	_____
Q 4. What was Mr. Bostick wearing?	_____	_____
Q 5. What was Mr. Bostick's reactions to being told that all he needed to do to appeal his conviction was to sign a form? (violent or non-violent)	_____	_____
Which slides did the subject see first? (stress or pleasant)	_____	

Comments:

APPENDIX C

Subjects Post-Interview Questionnaire

APPENDIX C

Michigan State University

Department of Communication

1. How successful do you feel you were in lying when you were questioned about the video-tape you saw before being interviewed?

extremely successful	1	2	3	4	5	6	7	8	9	10	extremely unsuccessful

Comments:

2. How successful do you feel you were when you lied about the slides you saw?

extremely successful	___	___	___	___	___	___	___	___	___	___	extremely unsuccessful

Comments:

3. How did you <u>really</u> feel when you saw the slides of burn victims?

Name _____

Age _____

Sex _____

Career Goals _____

_____

APPENDIX D

Guidelines for Subjects' Debriefing

APPENDIX D

1. Give them the questionnaire.

2. Most of what we told you today was true. However, we implied that the department of criminal justice was going to receive information about your performance on this task. This is false. In order to attempt to make this lying situation as involving as possible, we wanted you to think that your performance was somehow going to affect your law enforcement career. We are going to provide no information of any kind to the department of criminal justice about you. Nor will your name be associated in any way with data which this research provides or in any professional report of our findings.

3. The department of criminal justice is very interested in this research, as we said, and we are going to provide your name to Mr. Brandstedder indicating that you did participate (if you want). We really do appreciate your help.

4. We would like to show some short excerpts from some of the people we are interviewing here to some freshmen and sophomore students in the communication department next term and also to some adults in the area. We will ask them to try and identify the truthful responses. If we select some of your answers for this purpose, we will confidentially see to it that you get feedback as to how well they are able to do this in your case. We need to make it clear, however, that you may ask to have the tapes of your responses erased and we will do this, and it will in no way reflect negatively on you. Will you allow us to use your tapes? _____ Thank you. We really do appreciate your help.

5. We also have a favor to ask of you. Many students in criminal justice are participating in this exercise. So that others may have the opportunity to participate under realistic circumstances, it is important that they not know that the criminal justice department will not receive any information about their performance. So please do not talk to them about this research until after Thursday night.

6. If you have any questions, we'd rather not answer them until after Thursday night. You will be learning the full details of all aspects of this research soon.

APPENDIX E

Observer Questionnaire

APPENDIX E

Michigan State University
Department of Communication

We are conducting research which is funded by the National Science Foundation and is designed to examine the extent to which people can accurately detect deceptive communication, that is, lying. Earlier this year we created a situation in which some students who are planning careers in law enforcement were interviewed by a police officer. Each of the students had been instructed to lie to the officer at certain times and at other times to tell the truth. The situation we created was such that it was extremely important to the students that they successfully deceive the officer. They did not want the officer to be able to accurately detect their deception.

The students were videotaped during their interviews and today we would like you to watch these tapes and do your best to accurately detect when they were lying and when they were telling the truth. Participation in this research is voluntary. If you do choose to participate you will be seeing a total of 16 people engaging lying or "truthing" behaviors. We will provide feedback about how accurate you are in detecting the deception to everyone who wants this information. If you would like to receive your personal accuracy score, please put your name and a summer mailing address here:

Name _____

Address _____

_____

Instructor's Name _____

For each of the 16 people you will first see a short segment (about 30 seconds long) during which they always tell the truth. During this segment they were asked their name, major, year in school, hometown, and whether they had seen a recent football game. Immediately after this truthful segment, you will see another segment of the same person. This second segment will be the test segment and will show the person either lying or telling the truth. It is your task to determine which.

The first eight people you will see will be answering questions about a videotape they had seen. They saw one of several versions of this tape. On all versions a man is shown being sentenced to jail for a crime he had committed, but the nature of the crime, what the man is wearing, and his reactions to being sentenced are different. Those who lie about what was on the tape they saw were instructed to give answers which were consistent with a version of the tape which they did not see. Thus, the same answer could be truthful for one person and a lie for another.

The last eight people you will see will be lying or telling the truth about feelings they were experiencing as they watched some color slides on a television screen. They saw two separate sequences of slides. One set was very pleasant and included scenes such as lakes, boats, and children playing. The other sequence of slides showed individuals who had been seriously burned. As they watched these slides and immediately after they were over, they were asked questions about the mood or feelings which these slides created. They were instructed to say that they were having pleasant and happy feelings in answer to questions about both sets of slides. Of course, when they were watching the burn slides, their answers were lies.

The length of the segments you will see will vary and is unrelated to whether the person in the segment is telling the truth of lying. There will be a ten second pause between each person to give you time to mark this questionnaire indicating whether you think the person was lying or telling the truth during the test segment. Remember the first segment you will see of each person will always be truthful. Also, indicate how confident you are in your judgment by checking the space on the scale which most nearly reflects your degree of confidence in your answer.

The tape which you see might contain mostly deceptive segments, or mostly truthful segments, or about equal numbers of each. Please try to judge each sample without regard to your previous judgments. Once the tape begins, please do not talk to anyone. Make your own judgments. Your answers are anonymous and will never be associated with your name. If you do choose to participate please do your very best to accurately detect when an individual is being truthful and when he is lying. Thank you very much for your help. Your answers are very important to us and we appreciate your cooperation.

<u>We are first going to do a practice example</u>.

Example: Was the person in the practice example

lying ____ truthful _____

How confident are you in your judgment of whether this person was lying or telling the truth?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident

0 1 2 3 4 5 6 7 8 9 10

Remember, the first eight people you will see are lying or telling the truth about the videotape they had seen.

Segment 1

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 2

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 3

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 4

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 5

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 6

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 7

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 8

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 9

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 10

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 11

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 12

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 13

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 14

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 15

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

Segment 16

I think this person is: lying ___ truthful ___

How confident are you in your judgment?

not at all confident ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely confident
0 1 2 3 4 5 6 7 8 9 10

How good a liar do you consider yourself to be? That is, when you want to, are you able to successfully deceive:

your friends: never ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ always
0 1 2 3 4 5 6 7 8 9 10

strangers: never ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ always

How interesting or boring did you find the task of trying to detect deception?

extremely boring ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ extremely interesting
0 1 2 3 4 5 6 7 8 9 10

How reasonable a use of class time was this experiment?

very reasonable ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ very unreasonable
0 1 2 3 4 5 6 7 8 9 10

Were you sitting in a good position to observe this tape?

no, I couldn't see it well at all ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ yes, I could see it very well
0 1 2 3 4 5 6 7 8 9 10

We would now like to get some information about you.

Age _____

Sex _____

Grade point average _____

Race _____

Are you an American citizen? _____

Marital status _____

If you can tell us, on what basis did you make your judgments about whether a person was lying or telling the truth?

APPENDIX F

Observer's Results Letter

MICHIGAN STATE UNIVERSITY

College of Communication Arts
Department of Communication

East Lansing, Michigan 48824

August 27, 1975

Dear

This letter is in regard to the research on detecting deceptive communication which you participated in during your communication class last spring. The purpose of this study was to find out what categories of observable behavior people use to make accurate judgments of lying. You watched a videotape which showed 16 individuals either lying or telling the truth and attempted to accurately judge which was which. These videotapes were shown to a total of 730 communication students under 14 different conditions. Some groups observed a close-up of the head of the persons on the tape, others watched the head and body, and others saw the body only. Some observers both heard and viewed the individuals on the tapes, while others made their judgments based on the visual information only. Some students saw these tapes in color and others saw them in black and white. We put these three variables (Head only/Body only/Head and Body; Visual and Audio/Visual only; Color/Black and White) together in all possible combinations, thus creating 12 separate conditions. That is, there were different groups of students who watched these tapes under each combination of these conditions. So, for example, one group of observers viewed a black and white tape of only the body of those on the tapes, and also heard what was being said. Another group saw a close-up of the head in color, but did not have the sound, and so on. Finally, one group heard what the individuals being judged were saying but did not see them and one group read a written transcript of what they were saying. This resulted in the total of 14 different experimental conditions.

We have now compared the accuracy scores of groups of observers who saw the tapes under each of these conditions. In general the results indicated that those who saw the head and body were slightly more accurate than those who saw either the head or body only, who were about equally accurate. Observers who were provided with both the audio and visual information were quite a bit more accurate than those who were exposed to the visual material only. Those who saw the tapes in color were slightly more accurate than those who saw them in black and white. The accuracy score in each condition is shown below. Fifty percent represented what would be expected just by chance.

	Head Only		Body Only		Head & Body	
	Audio & Visual	Visual Only	Audio & Visual	Visual Only	Audio & Visual	Visual Only
Color	55.8%	48.4%	53.7%	48.4%	57.4%	48.5%
Black and White	52.6%	44.9%	53.0%	48.6%	54.0%	49.0%

Audio Only = 54.5% Transcript Only = 57.3%

Your personal accuracy score was _____ out of the 16 judgments you made. This task was very difficult and most people did not do better than what would be expected by chance (8 out of 16).

Your participation in this research project is very important for a variety of reasons. First, when the data you provided us are analyzed fully and written up for publication, our knowledge of the behavioral basis for accurate judgments of lying will have been advanced. This individual study, by itself, makes only a small contribution to our understanding of human communication behavior and its effects. However, other researchers interested in this same area will be able to examine our procedures and results and be able to extend and refine the knowledge gained. In this way, as more studies are completed, the accumulation of knowledge grows steadily and systematically.

Second, there are practical applications of this study. It was paid for by a National Science Foundation grant which is looking at the effects of using videotape during court proceedings. If, in the future, trials are recorded on videotape and shown to juries on television monitors, it will be important to know under what conditions jurors would have the best chance to detect lying by witnesses.

Third, the researchers have benefited from your participation in this research. It is necessary to conduct major research projects to complete graduate studies in all Social Science disciplines, including Communication. Your participation has been important to us personally for this reason.

Finally, you personally may have benefited from this experience. By participating in this experiment, by listening to the description of the research which was provided during the class period in which you participated, and by reading this letter, you may have learned a little bit about how Social Science research is conducted. Hopefully this project has been an educational experience for you. Also, much of the content of Social Science courses comes from research projects like this one. Thus, future communication students may benefit from your participation just as you have likely benefited from the contributions of past participants.

In short, your participation in this experiment was extremely valuable. We sincerely appreciate your help and would like to thank you very much.

Sincerely,

John Hocking
Graduate Assistant

APPENDIX G

Tables 17 and 18: Analysis of Variance for Factual and Emotional Accuracy Using the Transformed Scores

Table 17

ANOVA for Factual Accuracy* (using the transformed scores)

	Sum of Squares	DF	Mean Square	F	Probability Level
Main effects	14.423	5	2.885	18.309	.001
Colorbw	.222	1	.222	1.411	.233
Visaudio	10.099	1	10.099	64.103	.001
Tape	2.701	1	2.701	17.144	.001
Shot	1.907	2	.954	6.053	.003
2-way Interactions	4.240	9	.471	2.990	.002
Colorbw visaudio	.138	1	.138	.874	.999
Colorbw tape	.001	1	.001	.009	.999
Colorbw shot	.111	2	.056	.352	.999
Visaudio tape	.692	1	.692	4.390	.034
Visaudio shot	.074	2	.037	.233	.999
Tape shot	3.077	2	1.538	9.764	.001
3-way Interactions	1.611	7	.230	1.461	.178
Colorbw visaudio tape	.026	1	.026	.165	.999
Colorbw visaudio shot	.572	2	.286	1.814	.162
Colorbw tape shot	.017	2	.008	.052	.999
Visaudio tape shot	1.031	2	.516	3.273	.037
4-way Interactions	.341	2	.170	1.082	.340
Colorbw visaudio tape shot	.341	2	.170	1.082	.340
Residual	95.790	608	.158		
Total	116.405	631	.184		

*transformation formula: $a = 2(\sqrt{\text{arcsine } x}$

multiple $\underline{r}^2 = .124$

Table 18

ANOVA for Emotional Accuracy (using the transformed scores)

	Sum of Squares	DF	Mean Square	F	Probability Level
Main Effects	4.321	5	.864	8.228	.001
Colorbw	.056	1	.056	.530	.999
Visaudio	.154	1	.154	1.469	.224
Tape	3.148	1	3.148	29.975	.001
Shot	1.028	2	.514	4.895	.008
2-way Interactions	1.699	9	.189	1.797	.065
Colorbw visaudio	.130	1	.130	1.239	.265
Colorbw tape	.002	1	.002	.017	.999
Colorbw shot	.193	2	.096	.917	.999
Visaudio tape	.016	1	.016	.148	.999
Visaudio shot	.452	2	.226	2.154	.115
Tape shot	.862	2	.431	4.104	.017
3-way Interactions	.741	7	.106	1.008	.425
Colorbw visaudio tape	.003	1	.003	.027	.999
Colorbw visaudio shot	.205	2	.103	.977	.999
Colorbw tape shot	.033	2	.017	.159	.999
Visaudio tape shot	.525	2	.263	2.501	.081
4-way Interactions	.054	2	.027	.255	.999
Colorbw visaudio tape shot	.054	2	.027	.255	.999
Residual	63.854	608	.105		
Total	70.669	631	.112		

Multiple $\underline{r}^2$ = .061

APPENDIX H

Tables 19 and 20: Correlations Between Confidence and Accuracy for Both the Factual and Emotional Segments

Table 19

Correlations Between Accuracy and Confidence, and Probability Levels in Each Cell of the Design for the Factual Segments

		Head Only		Body Only		Head and Body			
		Visual & Aud.	Visual Only	Visual & Aud.	Visual Only	Visual & Aud.	Visual Only		
Color	I	.124 p=.277	-.058 p=.362	.131 p=.245	-.149 p=.288	.062 p=.365	-.277 p=.090	-.028	.021
	II	.007 p=.485	.134 p=.258	.118 p=.264	.351 p=.055	.141 p=.260	-.330 p=.058	.070	
Black & White	I	-.254 p=.096	.081 p=.374	.474 p=.013	-.108 p=.305	.123 p=.284	.143 p=.203	.077	.085
	II	.368 p=.035	-.172 p=.206	.044 p=.415	.056 p=.402	.042 p=.419	.218 p=.158	.093	
		+.061	-.004	+.192	+.038	+.092	-.062		
		.029		.115		.015			

Transcript Only

Tape		
	I	.037 p=.435
	II	.387 p=.034

Audio Only

Tape		
	I	-.254 p=.147
	II	.326 p=.076

Table 20

Correlations Between Accuracy and Confidence, and Probability Levels in Each Cell of the Design for the Emotional Segments

		Head Only		Body Only		Head and Body		
		Visual & Aud.	Visual Only	Visual & Aud.	Visyal Only	Visual & Aud.	Visual Only	
Color	I	.457 p=.012	-.160 p=.165	.326 p=.040	.080 p=.346	.239 p=.087	-.047 p=.410	.149
	II	.113 p=.291	-.156 p=.224	.128 p=.251	.242 p=.139	.078 p=.361	-.029 p=.447	.063
		.106						
Black & White	I	-.129 p=.257	-.216 p=.195	-.149 p=.259	.361 p=.038	.008 p=.485	.033 p=.425	-.015
	II	-.381 p=.033	.049 p=.410	.267 p=.094	.059 p=.395	-.045 p=.411	.320 p=.074	.045
		.015	-.121	.143	.186	.070	.069	
		-.053		.165		.070		
		.015						

Transcript Only

Tape	
I	.047 p=.417
II	.137 p=.272

Audio Only

Tape	
I	.200 p=.211
II	-.132 p=.285

REFERENCES

REFERENCES

Bem, D., and McConnell, H. Testing the self-perception explanation of dissonance phenomena: on the salience of pre-manipulation attitudes. Journal of Personality and Social Psychology, 1970, 14, 23-31.

Bermant, G. and Jacoubovitch, M. Fish out of water: A brief overview of social and psychological concerns about videotaped trials. The Hastings Law Review, 1975, 26, 999-1011.

Berrien, F. and Huntington, G. An exploratory study of pupillary responses during deception. Journal of Experimental Psychology, 1943, 32, 443-449.

Buck, R., Savin, V., Miller, R., and Caul, W. Communication of affect through facial expression in humans. Journal of Personality and Social Psychology, 1972, 23, 362-371.

Buck, R., Miller, R., and Caul, W. Sex, personality, and physiological variables in the communication of affect via facial expression. Journal of Personality and Social Psychology, 1974, 30, 587-596.

Cutrow, R., Parks, A., Lucas, N., and Thomas, K. The objective use of multiple physiological indices in the detection of deception. Psychophysiology, 1972, 9, 578-588.

Davis, R. Physiological responses as a means of evaluating information. In A. Biderman and H. Zimmer (eds.), The Manipulation of Human Behavior. New York: Wiley, 142-168, 1961.

Doret, D. Trial by videotape--Can justice be seen to be done? Temple Law Quarterly, 1974, 47, 228-268.

Eisenberg, A., and Smith, R. Nonverbal Communication. New York: Bobbs-Merrill, 1971.

Ekman, P. and Friesen, W. Nonverbal leakage and clues to deception. Psychiatry, 1969a, 32, 88-106.

Ekman, P. and Friesen, W. The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica, 1969b, 1, 49-98.

Ekman, P. and Friesen, W. Hand movements. Journal of Communication, 1972, 22, 353-374.

Ekman, P. and Friesen, W. Detecting deception from the body and face. Journal of Personality and Social Psychology, 1974, 29, 288-298.

English, H. Reaction-time symptoms of deception. American Journal of Psychology, 1926, 37, 428-429.

Fay, P. J. and Middleton, W. C. The ability to judge truth-telling or lying, from the voice as transmitted over a public address system. Journal of General Psychology, 1941, 24, 211-215.

Freedman, J. Role playing: psychology by consensus. Journal of Personality and Social Psychology, 1969, 13, 107-114.

Goldstein, E. R. Reaction times and the consciousness of deception. American Journal of Psychology, 1923, 34, 562-581.

Gustafson, L. and Orne, M. Effects of heightened motivation on the detection of deception. Journal of Applied Psychology, 1963, 47, 408-411.

Hays, W. Statistics. New York: Holt, Rinehart, & Winston, 1963.

Hildreth, R. An experimental study of audiences' ability to distinguish between sincere and insincere speeches. Unpublished doctoral dissertation, University of Southern California, 1953.

Hocking, J., Margreiter, D., and Hylton, C. Intra-audience effects: A field test. Human Communication Research, in press.

Horvath, F. Verbal and nonverbal clues to truth and deception during polygraph examinations. Journal of Police Science and Administration, 1973, 1, 138-152.

Katzman, N. I. Violence and color television: What children of different ages learn. Unpublished manuscript, Department of Communication, Michigan State University, 1971.

Katzman, N. I. and Nyenhuis, J. Color versus black and white effects on learning, opinion, and attention. Unpublished manuscript, Department of Communication, Michigan State University, 1971.

Kerlinger, F. and Pedhazur, E. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1973.

Knapp, M. Nonverbal Communication in Human Interaction. New York: Holt, Rinehart and Winston, Inc., 1972.

Knapp, M., Hart, R., and Dennis, H. An exploration of deception as a communication construct. Human Communication Research, 1974, 1, 15-29.

Lee, C. The Instrumental Detection of Deception: The Lie Test. Springfield, Illinois: Thomas, 1953.

Maier, N. and Thurber, J. Accuracy of judgments of deception when an interview is watched, heard, and read. Personnel Psychology, 1968, 21, 23-30.

Maier, N. Sensitivity to attempts at deception in an interview situation. Personnel Psychology, 1965, 19, 55-65.

Maier, N. and Janzen, J. The reliability of reasons used in making judgments of honesty and dishonesty. Perceptual and Motor Skills, 1967, 25, 141-151.

Matarazzo, J., Wiens, A., Jackson, R., and Manaugh, T. Interviewer speech behavior under conditions of engogenously-present and exogenously-induced motivation states. Journal of Clinical Psychology, 1970, 26, 141-148.

Marston, W. Reaction-time symptoms of deception. Journal of Experimental Psychology, 1920, 3, 72-87.

McClintock, C. and Hunt, R. Nonverbal indicators of affect and deception in an interview setting. Journal of Applied Psychology, 1975, 5, 54-67.

McCrystal, J. Ohio's first videotape trial. The Ohio Bar, 1972, 45, 1-4.

Mehrabian, A. Communication without words. Psychology Today, 1968, 2, 51-52.

Mehrabian, A. Nonverbal betrayal of feelings. Journal of Experimental Research in Personality, 1971, 5, 64-75.

Miller, G. and Burgoon, M. New Techniques of Persuasion. New York: Harper and Row, 1973.

Morrill, A. Trial Diplomacy. Chicago: Court Practice Institute, 1971.

Motley, M. Accoustic correlates of lies. Western Speech, 1974, 37, 81-87.

Nie, N., Hull, C., Jenkins, J., Steinbrenner, K., and Bent, D. Statistical Package for the Social Sciences. New York: McGraw-Hill, 1975.

Scheffé, H. A method for judging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Shulman, G. An experimental study of the effects of receiver sex, communicator sex, and warning on the ability of receivers to detect deceptive communicators. Unpublished master's thesis, Purdue University, 1973.

Trovillo, P. A history of lie detection. Journal of Criminal Law and Criminology, 1939, 29, 848-881.

Winer, B. Statistical Principles in Experimental Design. New York: McGraw-Hill Co., 1971.

Wolk, R. and Henley, A. The Right to Lie. New York: Wyden, 1970.