



This is to certify that the
dissertation entitled

Accuracy In Deception Detection:

A Quantitative Review
presented by

Pamela J. Kalbfleisch

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Communication


Major professor

Date August 1, 1985



RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

~~JUN 21 1977~~

~~177 A135~~

~~SEP 21 1977~~

~~000 H310~~

~~MAY 11 1982~~

~~22 012~~

~~050~~

~~DEC 0 8 1993~~

~~7117264~~

~~MAY 26 1997~~

~~SEP 16 1998~~

~~4850313~~

SEP 21 2001
1123 01

JAN 07 2002

MAY 01 2002 2

ACCURACY IN DECEPTION DETECTION:

A QUANTITATIVE REVIEW

By

Pamela Joy Kalbfleisch

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Communication

1985

ABSTRACT

ACCURACY IN DECEPTION DETECTION: A QUANTITATIVE REVIEW

By

Pamela Joy Kalbfleisch

This dissertation reports a meta-analysis of the research on the human ability to detect deception. A review of literature and critique of previous deception detection meta-analyses are considered prior to the development of a rationale for the meta-analytic technique designed for this analysis. The predominant use of repeated measures and mixed measures research designs in the deception research negated the use of traditional meta-analytic techniques, because these techniques do not feature a method to extract effect sizes or to cumulate the results of these designs. The meta-analytic technique developed for this meta-analysis allows the researcher to cumulate the results from dichotomous measurements in repeated measures, mixed measures and between-subjects designs, but does not estimate sampling error, measurement error nor restriction in range.

In general, this meta-analysis found humans to be poor lie detectors. Differences in this ability across experimental conditions are small. However, several patterns of human success in deception detection are apparent. Specifically, the presence

Pamela Joy Kalbfleisch

of message content allows observers to obtain higher accuracy scores than in observation conditions without content. Observers viewing shots of communicators' bodies are more accurate than observers viewing combination shots of heads and bodies, or shots of heads only. Evidence also suggests observers are more likely to judge a communicator to be telling the truth than to be lying. Other suggested patterns were 1) observers familiar with the truthful behavior of the person they are judging are more accurate than those not exposed to a truthful baseline, 2) females appear to be slightly more accurate than males in detection deception, and 3) female deceivers appear to be slightly easier to detect than male deceivers.

Dedicated to my parents
Paul and Marian Kalbfleisch

ACKNOWLEDGMENTS

It has been a privilege to observe and work with my advisor Dr. Gerald R. Miller during my years at Michigan State. With his guidance, I have learned much about scholarship and professionalism. A hearty thanks G.R.!

Dr. John E. Hunter contributed to my training by sharing his creations of research methodologies for the future. It has been exciting to be aware of his mathematical innovations years before they will be studied in most college classrooms.

Dr. William A. Donohue always provided helpful advice. I have often been grateful for his savvy and perceptive observations. Whether such understanding comes naturally or is the result of spending years studying talk, is not clear to me.

Dr. Raymond Frankmann is responsible for much of my statistical knowledge. To my surprise during the defense of this dissertation, he emerged as a mechanical lie detection expert, having spent numerous hours administering polygraph examinations.

Dr. Charles Atkin helped me complete my graduate work with his input on this project. I appreciate having the opportunity to include this communication scholar on my guidance committee.

Dr. Edward Fink, although he was not part of my guidance committee, contributed significantly to my graduate training. Since leaving Michigan State, he has continued to provide unwavering support of my research and professional development.

James Stiff and Paul Mongeau were trusted research colleagues. I will always appreciate their friendship.

Finally, those closest to me deserve the biggest thanks of all. My parents Paul and Marian Kalbfleisch have provided years of support for me regardless of the venture. From the rough and tumble rodeo days to my scholarly pursuits. They have never lost faith in me, nor let me lose faith in myself. To them I owe much.

My dear friend Jan David Gierman provided hours, months and years of understanding and help. He has been a positive force in my life. Having never known me other than as a graduate student, I know he is eagerly anticipating "life without graduate school".

My Sister Karen Lind and her husband Eugene, were also supportive through this endeavor. I am pleased to have been blessed with such a sibling and brother-in-law.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
INTRODUCTION	1
Overview	1
Lie Detection in the Past: The Ordeals	2
Modern Physiological Detection of Deception	3
Social Science Examination of Unaided Human Human Ability to Detect Deception	7
The Significance Test Tallies	11
The Quantitative Reviews	13
Zuckerman, DePaulo and Rosenthal	14
Kraut	18
The Proposed Meta-Analysis	24
METHOD	25
Extraction of Effect Sizes for Repeated Measures: The Special Case	27
Sample Selection	31
Procedures for Cumulation	39
RESULTS	41
Observational Conditions	41
Familiarity	49
Sex Differences in Deception Detection	51
Judgments of Truth vs Lie	56
Narrative Review	57

DISCUSSION		67
Comparison With Previous Meta-Analyses		67
Sampling Error		71
Conclusion		72
REFERENCES		74
Footnotes		91

LIST OF TABLES

Table	Page
1. Studies Selected for Meta-Analysis	33
2. Deception Detection Accuracy: Conditions Under Which Truthful and Deceptive Messages are Observed	43
3. Deception Detection Accuracy: Conditions Under Which Truthful and Deceptive Messages are Observed	45
4. Accuracy in Deception Detection: Conditions Under Which Truthful and Deceitful Messages are Observed	48
5. Baseline Familiarity With Communicator and Judges Accuracy in Deception Detection	50
6. Judges Sex and Accuracy in Deception Detection	52
7. Deception Detection Accuracy for Male Judges	53
8. Deception Detection Accuracy for Female Judges	53
9. Male Judges: Deception Detection Accuracy Observing Male vs Female Communicators	54
10. Female Judges: Deception Detection Accuracy Observing Male vs Female Communicators	55
11. Sex of Communicators and Accuracy in Deception Detection	56
12. Observer Judgments of Truth vs Lie	57

INTRODUCTION

Overview

There has long been concern over the accuracy of decisions indicting suspected criminals and untrustworthy social contacts. History is marked with attempts to ward off uncertainty associated with assessments of veracity. The last eighty years have seen American law enforcement and businesses turn to devices such as the polygraph in order to avoid the uncertainty involved in lie detection.

Social scientists have tried to assess human abilities without the aid of such devices. Results of these investigations have been both contradictory and complex. Through the use of quantitative reviews, Zuckerman, DePaulo and Rosenthal (1981) and Kraut (1980) have attempted to make this body of research easier to comprehend and to provide overall assessments of human lie detection ability. However, these attempts have shortcomings.

This dissertation first briefly reviews the lie detection literature beginning with the techniques which have attempted to substitute a tool or device for human judgment, followed by a look at the social science research concerning human lie detectors. Then each quantitative review of the deception detection research is examined in depth followed by a

reexamination of accuracy in deception detection using an alternate analytical technique.

Lie Detection In The Past: The Ordeals

Uncertainty over possible lies has long been a troubling problem for members of the human race. As early as 900 B.C. writings on papyrus outlined steps to detect liars (Trovillo, 1939a, 1939b).

Attempts to avoid relying on human judgment began with the use of physical ordeals to ferret out liars (Trovillo, 1939a, 1939b). The ordeals assigned the decision making to forces the society perceived as being greater than themselves (i.e. the gods). Trovillo (1939a, 1939b) reviewed techniques such as the red-hot iron ordeal where accused liars were forced to lick red-hot irons, being exonerated if they emerged unburned, or the balance ordeal, where the suspected were weighed before and after a mystical exhortation with proof of veracity shown by a change in the weight of the person under suspicion.

Trials by ordeal were still used in the third world countries at the time Trovillo published his review. For example, in some villages in India, those suspected of deception were made to chew rice, guilt being determined if rice could not be swallowed. In Africa deceit was determined by plunging one's arm in boiling water and then looking for tell-tale "deception blisters" the next day.

Some of these procedures may seem primitive. However, the African medicine man sniffing potential deceivers in order to ferret out deception may not be much different from the physiological measures of palmer sweat used by modern mechanical lie detectors. The dry mouth unable to swallow rice and the dry tongue surface seared by the hot iron may also be related to the physiological phenomena of "dry mouth" or decreased saliva flow during times of high stress. Kleinmuntz and Szucko (1984a) in their examination of primitive lie detection techniques noted that if the mouth of the accused remained moist it could have been possible to avoid being burned by the iron and possible to swallow the dry rice.

The accuracy of these primitive lie detection techniques remains a mystery. The decreased flow of saliva or the increased sweating accompanying fear of discovery may well have accompanied fear of the ordeal itself. Also the physiological processes underlying ordeals such as plunging one's hand into boiling water and weight change on a balancing scale still remain unexplained and their relationship to actual lie detection is vague.

Modern Physiological Detection of Deception

Modern attempts to detect deception have also focused on physiological signs: breathing change (Benusi, 1914), variation in systolic blood pressure (Marston, 1917; Larson, 1932) and

galvanic skin responses (Summers, 1937, 1939). Modern polygraphs now measure changes in all three of these physical processes (Yohman, 1978).

The mechanical lie detector has had a substantial impact on the American public. Lykken (1981) estimated that over one million Americans were subjected to polygraph tests in 1980 either in the legal system or employment screening. Kleinmuntz and Szucko (1984a) noted that the American Polygraph Association projected 2.3 million Americans would be taking polygraph tests in 1984. However, despite increasing use of the technique, the "ordeal" of the polygraph remains a controversial issue. The controversy centers on mixed empirical evidence regarding the accuracy of the polygraph and problems in validity of measures of the machine's accuracy.

Mixed empirical evidence. In 1977 Podlesny and Raskin published results indicating that the polygraph was accurate from 88 to 96 percent of the time. In 1978 Raskin and Hare reported findings of 96 percent accuracy and 99 percent accuracy in a subsequent study (Raskin, Barland & Podlesny as cited in Raskin, 1978). Raskin and Podlesny (1979) broke down the results of their lie detector tests yielding findings of 90 percent accuracy levels when subjects were lying and 89 percent accuracy when they were telling the truth.

However, Lykken (1979) failed to find the high accuracy ratings reported by these researchers. Instead he found the

polygraph to be 64 to 71 percent accurate with a strong bias against truthful communicators: 36 to 39 percent of the truthful communicators were incorrectly classified as deceptive. Kleinmuntz and Szucko (1984b) and Szucko & Kleinmuntz (1981) found the false positive rate (number of truth-tellers classified as deceptive) to be as high as 55 percent.

Measures of polygraph accuracy. Lykken (1978, 1979) and Kleinmuntz and Szucko (1984a) have questioned the use of the polygraph in lie detection claiming the accuracy of the machine has not sufficiently been established and the tests of the technique are fraught with problems. These researchers note that although a technique for interpretation of the results does exist in which fluctuations in the polygraphs records are submitted to a set of measurements and scoring (Backster, 1963), many interpreters eschew specific measurements and simply make global assessments of communicators' veracity based on their overall performance. After analyzing polygraph records, Kleinmuntz and Szucko (1981) concluded that the interpreters' judgments often bore little resemblance to the physiological data.

While laboratory studies may carefully control the interpreter's knowledge of the experimental conditions, the interpreter's impressions of the communicator remain unchecked. Field studies also share this problem, Ben-Shakhar, Lieblich and Bar-Hillel (1982) contend that very few interpreters use any tec-

hunique other than global evaluations in assessing veracity. Also, there may be contamination of lie detector results because additional information about the suspect may be at the interpreter's disposal, e.g. intelligence gathered about the suspect, impressions of former interrogators, and records or previous convictions (Ben-Shakhar et al., 1982).

The polygraph accuracy studies done in field settings also lack a verifiable criterion of correctness. Accuracy in criminal or civil legal cases is assessed either by comparing the polygraph test results with the decisions of the judge, jury or panel of legal experts, or by a confession of the accused (Barland & Raskin, 1976; Bersh, 1969; Horvath, 1977; Horvath & Reid, 1971; Hunter & Ash, 1973; Slowik & Buckley, 1975; Wicklander & Hunter, 1975). Thus, whether the polygraph is actually accurate in these settings is verified only through human judgment.

Modern lie detection appears to depend upon the human ability to detect deception. Members of the Reid and Arther schools of lie detection (Reid & Inbau, 1977) apparently are proud of this fact. Lykken (1979) noted numerous endorsements in the polygraph trade journals for global assessments of suspected deceivers, stressing that it was the technician who administers and interprets the test and the technician who, in fact, functions as the lie detector and not the polygraph.

In a response to Lykken's (1978) critique of Raskin and Hare's (1978) high polygraph accuracy ratings, Raskin (1978)

disputed Lykken's criticism that one of the reasons for the high ratings could be that the interpreters were very well informed and free to make clinical interpretations. Raskin (1978) defended his findings by asserting the human factor in lie detection decisions is negligible. His defense of the polygraph's high accuracy scores was based on the assumption that humans are very poor lie detectors, an assumption that appears to underlie use of alternative lie detection methods. However, the human impact of the truth/lie classifications of the polygraph is hard to assess. In field studies high polygraph scores represent concurrent agreement between human and machine, and not a measure of accuracy. Additionally, in both lab and field studies the polygraph accuracy is the result of the combination of human and machine classification.

Social Science Examination of Unaided Human Ability to Detect Deception

Social scientists have studied the lie detection abilities of human judgment without the polygraph. Studies as early as 1941 (Fay & Middleton) have probed the question of how accurate humans might be in their attempts to tell lies from truths. Review of these findings is important to determine the degree of accuracy that could be expected in legal setting by judges, jurors or polygraph interpreters. It is also important to see how well humans may cope with the everyday social milieu.

The soundness of decisions made regarding whether to believe a tardy lover's excuse, a repairman's cost estimate, or an acquaintance's statement that a wallet left at a dinner table was never noticed or found relies on an individual's own lie detection ability. Friendships may end because a friend's truthful statements were doubted; individuals may be exploited by false claims, and frustration may build as one wonders just whom to believe when conflicts arise. While these decisions may not be life and death matters, everyday veracity assessments may affect the course of human relationships and the ability to manage one's environment.

Contradictory findings. Studies of human lie detection have explored the effects of a variety of observational conditions on human accuracy, investigated an assortment of observers and their differential abilities and examined the deceptive performance of differing types of communicators to see if the falsehoods of some are more difficult for observers to detect than others. The results of these studies appear complex and often contradictory.¹ For example, while human accuracy at detecting lies and truths has sometimes been found to be significantly better than chance (Hemsley & Doob, 1979; Lavrakas & Maier, 1979; Potamkin, 1982); others have found these accuracy levels do not exceed chance expectations (Bauchner, Brandt & Miller, 1977; Matarazzo, Wiens, Jackson & Manaugh, 1970; Motley,

1974). The mediating effects of observational conditions also remain unclear. Ekman and Friesen (1974) theorized that deceptive communication could be detected better by watching the bodies of potential deceivers rather than their heads and faces. Ekman and Friesen posited that people are more aware of their faces and hence exert more control over them both when deceiving and truth-telling. This awareness allows them to use their faces to more accurately convey messages and to more carefully mask deception clues. Conversely, Ekman and Friesen reasoned that people are less aware of their bodies and thus exert less control over body cues that might mark their deceptive attempts. These researchers back up their "leakage hypothesis" with empirical results indicating observers who viewed only the bodies of communicators made more accurate truth/lie decisions than those viewing the heads.

Littlepage and Pineault's (1981) findings are consistent with those of Ekman and Friesen (1974). When communicators express emotional information, such as descriptions of moods and feelings, Hocking, Bauchner, Kaminski and Miller's (1979) research is also consistent with the Ekman and Friesen theory. However, when communicators lied and told the truth regarding factual information, Hocking et al. (1979) found observers viewing communicators' heads were more accurate than those viewing shots of their bodies. Finally, Wilson (1975) found

neither the head nor the body, but rather the combination of both, was the best for accurate judgments of veracity.

Maier and Thurber (1968) and Sakai (1981) removed the viewing condition altogether. Their results indicate that those listening to communicators were better able to accurately discriminate lies from truths than those viewing them. These results were not found in other studies. For example, Wilson (1975) noted that those who listened to deceptive and truthful communication were not as accurate as those who viewed the communication. Bauchner et al. (1977), Hocking et al. (1979) and Harrison et al. (1978) found no significant differences in judgmental decisions between the audio and video conditions.

Deception studies have not established whether the sexes differ in gullibility. Sakai (1981) and Maier and Thurber (1968) contend that females are better lie detectors than males. However Atmiyanandana (1976), Parker (1978) and Rovira (1982) conclude there is no difference between the sexes. Parker (1978) noted that while a sex difference in observers was not found, observers did have more difficulty detecting the deceptive communication of female than of male deceivers. Sakai (1981) also noted a significant difference in success at detecting male versus female deceivers; however, Rovira (1982) and Wilson (1975) found no significant differences in detection of male versus female deceivers.

These studies illustrate the contradictory results in the deception literature. The findings do not yield clear statements regarding how accurately humans can detect lies from truths or how observational conditions and individual difference variables may mediate this accuracy.

The Significance Test Tallies

The divergent findings of the deception research might prompt a reviewer to start a tally of significant and nonsignificant results, with the goal of allowing a preponderance of evidence, either significant or nonsignificant, to arbitrate these mixed findings. In the literature reviews that open most deception studies this tallying has implicitly taken place. In rationales for research questions and hypotheses, researchers often depend on significance tests reported by other researchers to build a line of research reasoning. This line of reasoning may indicate what questions have yet to be answered, propose the presence of possible moderating variables or simply report past research that supports the current investigation.

After reviewing the diverse findings in deception research, a number of social scientists have called for more research to assess human lie detection abilities (Feldman & White, 1980; Harrison et al., 1978; Hocking et al., 1979; Hemsley & Doob, 1979; Maier & Lavrakas, 1976; Parker, 1978; Rovira, 1982; Rotkin, 1980; Sereno, 1981). In light of the numerous completed

deception studies, this plea overlooks the possibility that additional exploration will only add to the contradictions. Other researchers (Comadena, 1982; Potamkin, 1982) have attempted to explain the contradictory findings by pointing out methodological limitations in those studies that contradict their line of reasoning.

Problems in using significance tests for cumulation.

Unfortunately, the use of significance test tallies is not the best approach for better understanding the findings dealing with human lie detection. Attempting to cumulate results across studies by counting significant findings versus nonsignificant findings can yield misleading results.

One problem is the significance test does not provide an estimation of the strength or importance of a relationship. It only suggests whether the obtained F differs significantly from chance (Glass, McGaw & Smith, 1981). When cumulating significance tests, significant results are classified as presence of a relationship and nonsignificant results are classified as absence of a relationship (Hunter, Schmidt & Jackson, 1982). Confidence intervals for obtained values are not considered in these tallies. Hence, significant and nonsignificant results are classified as absolute (Glass et al., 1981). Further, since results from large sample sizes are more likely to be found significant than those from small samples,

some strong effects from small samples may be overlooked (Glass et al., 1981).

The Quantitative Reviews

Beyond the significance test tallies, two recent attempts have been made to quantitatively review the deception research and provide an answer to some of the unanswered questions concerning lie detection. Zuckerman, DePaulo and Rosenthal (1981) and Kraut (1980) used meta-analysis to evaluate research findings. The large number of subjects in these aggregated analyses should allow researchers to better estimate the presence of actual effects (Hunter, 1982). Glass et al. (1981), Hunter et al. (1982), and Levine, Romashko and Fleishman (1973) have all developed techniques that use summary statistics other than the significance test to cumulate study findings and thus avoid the test's dependence on sample size. Hunter et al. (1982) have also developed methods to assess the degree of sampling error present in the cumulated data along with other statistical artifacts such as measurement error and restriction in range (see Hunter et al. 1982 for a review of these meta-analytic techniques and others).

The quantitative reviews of the deception detection literature by Zuckerman et al. (1981), and by Kraut (1980) use cumulative methods that differ from those developed by Glass et al. (1981), Hunter et al. (1982) and Levine et al. (1973). These deception meta-analyses will be reviewed in terms of their

usefulness, application, and contribution toward the understanding of human lie detection.

Zuckerman, DePaulo and Rosenthal

The most extensive of these quantitative reviews was completed by Zuckerman et al. (1981). This project reviewed both the literature concerning accuracy in deception detection and that which correlated verbal and nonverbal behaviors with deceptive communication. They sought to use Cohen's d statistic as a measure of effect size. They computed this statistic for this review as $\frac{2\sqrt{F}}{\sqrt{df}}$ where F equals the obtained value for F and df equals the degrees of freedom for the variable of interest. However, this computation is inappropriate for studies with repeated measures, i.e. the majority of deception studies.

Conceptual issues. In Zuckerman et al. (1979), studies are presented with their corresponding d estimate of effect size. The problem is that d as defined is a statistic based on a comparison (Cohen, 1977). However, only one set of comparisons made in the meta-analysis is defined. In this set the d is the effect size of the difference between truthful and deceptive judgments. In the other sets of comparisons it is unclear exactly what is being compared. In interpreting the d values provided, knowledge of what comparisons the effect estimates represent is critical.

A second problem in interpreting the tables in Zuckerman et al. (1979) is that the results are provided in standard deviation units and no other summary information is made available for interpreting these findings. The standard deviations alone do not provide sufficient information from which to determine how accurate humans are at lie detection. For example, how accurate are people at detecting deception at 1.2 standard deviations or at 1.4? The standard deviations suggest that the observers in one study are more accurate than those in another, but the magnitude of this accuracy is not specified.

Operational issues. While these are both problems that affect interpretation of the results of the Zuckerman et al. (1979) study, the inappropriate computation of Cohen's d statistic in the estimation of effect sizes is far more serious. Cohen's d statistic as computed by Zuckerman et al. and applied to the deception research may have yielded estimates which are grossly over-inflated, leaving the overall validity of the Zuckerman et al meta-analysis in question.

The d statistic as defined in this meta-analysis is actually a transformation of F to t with the sample size removed from the t to obtain d . The df in the denominator is used for the estimate of sample size. The trouble with this effect size statistic is that it is not appropriate for the most common designs used in deception research. The d statistic described by

Zuckerman et al. assumes an independent groups analysis of variance design with two levels.

For example, the use of degrees of freedom as an estimate of sample size is inappropriate in the case where multiple factors are used in a study. In the independent group designs where the degrees of freedom equal $N-1$, use of the degrees of freedom to estimate sample size of this value is equal to the sample size minus one. However, in the multiple factor independent group designs the degrees of freedom will be $N-K$, where K equals the number of factors in the design. This error in Zuckerman et al.'s application of the d statistic to multiple factor designs will yield d 's that are artificially large due to the reduced size of the denominator with $N-K$ replacing the $N-1$ for multi-factor designs.

While this sample size estimate is a small problem, a more serious error is inherent in applying the d statistic to research using the research design with repeated measures. These designs are implemented in deception detection by having each observer judge the veracity of a number of communicators who lie or tell the truth. Typically each judge will view a number of communicators who represent different manipulations of the independent variables in the design. The judge may also be asked to make the decisions in different viewing conditions. The differences in judges are typically treated as between-subjects factors and the differences in the communicators and the

conditions under which they are observed then are treated as within-subjects factors (repeated measures).

The major problem with the Zuckerman et al. use of the d statistic is that it fails to properly estimate the effect size of these repeated measures. There are several reasons for this failure. The first problem with estimating the effects of within-subjects factors in this way is that the variance of within-subjects factors is the variance of difference scores and not an estimate of the within-cell variance. The d as used by Zuckerman et al. assumes an estimate of the within-cell variance. Therefore it does not properly assess the variance of difference scores.

The degrees of freedom estimate for the within subjects F values are also not comparable to the degrees of freedom value assumed by Zuckerman et al. to be $N-1$. Within-subjects F values will differ depending on the error term associated with the degrees of freedom. This difference in the size of the degrees of freedom is another reason the Zuckerman d values estimated from within-subjects designs will not be comparable to those from between-subjects designs.

These problems with the d statistic in the within-subjects case created by an inappropriate estimate of variance will manifest themselves in over inflated d values. Also, Zuckerman et al.'s meta-analytic estimation of effect size based on the

value of F for within-subjects factors may not even be possible without access to the original data. Current statistical reporting process in many social science journals that publish deception research does not include the publication of the variance estimates necessary for determining the effect sizes for these within-subjects designs.

A final criticism of the Zuckerman et al. meta-analysis is that for singular studies, estimates of effect size were not provided. Instead, Zuckerman et al. dealt with these studies by narratively describing their results in terms of statistical significance. Since the authors of these studies provided the necessary summary statistics for the computation of Zuckerman's d statistic, the results of the individual studies should have been provided in a common metric with the studies cumulated in the meta-analysis.

The Zuckerman et al. meta-analysis is the most comprehensive quantitative review of the deception literature to date. However, it has severe problems in the operationalization and application of the Cohen's d effect size statistic and hence the results should be considered with caution.

Kraut

The Kraut (1980) quantitative literature summary of accuracy in deception detection is less extensive. He presents a fairly simple analysis of this research in the form of an average compiled from the results of ten studies. His estimate of average

human lie detection ability is 57 percent where 50 percent accuracy would be expected by chance.

The summary statistic used by Kraut, the mean, is easier understood than the Zuckerman et al. effect size measures. For example, if observers are reported as being able to tell lies from truths 50 percent of the time, it is easy to picture observers being incorrect in half of their judgments. A 75 percent accuracy level can easily be interpreted as being incorrect in one out of four judgments or being correct three fourths of the time.

However, Kraut's results may be misleading. In Kraut's review he summarizes a sizable body of complex and somewhat contradictory studies with a single average (57%) and its standard deviation (7.8%). Overlooked is the fact that these studies have measured accuracy of judges placed in several different types of observational conditions in experiments testing the effects of a number of different independent variables in varying experimental paradigms.

For example, Kraut and Poe (1980) looked at the accuracy of customs inspectors compared to laypersons, while Hemsley (1977) tested male versus female ability to detect deceptive communication of members of the same and opposite sex. Maier and Thurber (1968) had their communicators role play the parts of deceivers or truth-tellers while observers watched, heard or read

the transcript of their performance, Ekman and Friesen (1974), on the other hand, had their communicators respond deceitfully or truthfully to an interviewer's questions. Observers in the Ekman and Friesen study viewed silent video tapes of these responses, but were allowed to view only the heads or bodies of the stimulus persons. Kraut's single overall average does not indicate what differences if any there may be under these different conditions and others.

By consolidating the accuracy ratings for each study into one score for averaging, Kraut may have misrepresented the findings of this body of research. For example, Maier and Thurber (1968) found that those who listened to and those who read transcripts of people lying and telling the truth were more accurate (77.0% and 77.3%) than those who watched people lying and telling the truth (58.3%). However Kraut averaged these conditions to yield 70.9 percent as the representative summary statistic for the Maier and Thurber study in his overall cumulation of the deception studies. By doing this Kraut not only overlooked differences in experimental conditions when constructing his overall average but his results may also be misleading, with an overall average of a study's accuracy levels across conditions not providing a very good picture of the outcomes of the Maier and Thurber study or others.

Further problems in the utility of Kraut's overall estimate lie in the sample of studies he selected for inclusion in the

average. These problems lie in the: 1) derivation of mean accuracy ratings, 2) chance probability associated with included estimates, and 3) the selection procedures used in obtaining the mean accuracy ratings.

If one reviews the published reports of the studies included in Kraut's overall estimate of accuracy several of these problems become apparent. For example, Kraut utilized an accuracy estimate of 46 percent from the Kraut and Poe (1980) experiment. However, their published report does not contain any reference to a mean accuracy estimate. Possibly Kraut's use of an unreported mean estimate can be understood given he was senior author of the study and had ready access to the unpublished raw data.

This was not the case in the Maier and Janzen (1967) article. In this report, counts were provided, recording the number of observers who correctly identified liars and truth-tellers. To obtain a mean accuracy rating in a 0-1 metric, Kraut was required to convert this count to a percentage by dividing the number of correct judgments by the total number of judgments made. Dividing the Maier and Janzen count of 97 correct judgments by the 228 total judgments yields a mean accuracy estimate of 43 percent. However, Kraut listed an estimate of 61 percent from the Maier and Janzen study in constructing his overall mean accuracy.

The origin of the mean accuracy estimates that Kraut used to represent the Kraut and Poe (1980) and Maier and Janzen (1967) studies is unclear. A brief explanation by Kraut regarding their derivation would have added credibility to their usage as estimates of accuracy in deception detection.

A second problem in the sample of mean accuracy ratings employed by Kraut lies in the chance level of one of the studies. While Kraut interprets his overall estimate as having a chance level of 50 percent in all studies, the chance level in the Geizer, Rarick & Soldow (1977) study appears to be 33.3 percent. Instead of asking observers to decide whether people were lying or telling the truth (where chance would be 50%), they asked their observers to decide which of three people was telling the truth in a "To Tell The Truth" game show excerpt. In this case the possibility that the judges would make the correct choice by chance was 33.3 percent. In Kraut's article he notes this difference in probability, but he still includes the Geizer et al. estimate in his overall average of detection ability. In doing this he has created an overall estimate based on estimates not converted to the same metric, hence the Hemsley and Doob estimate is not comparable to the others included in the same overall estimate.

The specific sample of studies may also contribute to the problems with Kraut's review. The total number of lie detection judges is 855 people. Across 32 studies² reporting mean accuracy

ratings in a metric ranging from 0 to 1 (with .5 chance accuracy level), there are 3439 possible lie detection judges in the deception literature or 3,577 if one wishes to include the Kraut and Poe (1980) and Geizer et al. (1977) studies. Why Kraut choose to use less than half the available studies is not clear, nor are the selection procedures that he used to compile his sample. For example, Kraut chose to use three of the four deception detection studies reported by Maier and his associates (Lavrakas & Maier, 1979; Maier & Janzen, 1967; Maier & Thurber, 1968), but did not indicate why the Maier and Lavrakas (1976) study was not included. Other lines of research were totally excluded from the quantitative review. This incomplete selection of studies brings the representativeness of the Kraut estimate of human lie detection ability into question.

In summary, problems with the quantitative review by Kraut include use of a small unrepresentative sample, failure to differentiate the variables considered in studies included, and utilization of misleading accuracy estimates that resulted from study-wide means pulled into the middle range by averaging overall experimental conditions. Use of estimates that were possibly unsuitable for this review also added to the reduced utility of this quantitative summary.

The Proposed Meta-Analysis

Critiques of the Zuckerman et al. (1979) and Kraut (1980) meta-analyses indicate that the results yielded by these analyses are misleading. An accurate assessment of human lie detection ability and those observational conditions and individual difference variables that may mediate it are still not available.

Given our dependence on deception detection ability in legal, business and social settings, an assessment of this skill is essential. A reexamination of the deception detection literature in a different context should provide a more enlightening understanding of our ability to detect deceit. Because this body of research relies heavily on mixed and repeated measures designs, this dissertation will present a meta-analysis that uses a technique allowing for cumulation of effect sizes across these designs. The meta-analytic techniques of Hunter et al. (1982), Glass et al. (1981) and Levine et al. (1973) currently do not provide the methods necessary for this type of cumulation.

In conjunction with the presentation of this meta-analysis the cumulation technique will be explained as will the strategy for study selection suitable for use with this method. This explication will be followed by the results of the meta-analytic procedure and a discussion of these findings.

METHOD

The deception detection research with its habitual choice of the within-subjects and mixed research designs, does not easily lend itself to traditional meta-analytic procedures. The most widely used meta-analytic techniques were developed for designs that utilize between-subjects factors (Glass et al, 1981; Hunter et al, 1982; Levine et al, 1978; Rosenthal, 1978).

The reason that the effect sizes have not been assessed for within-subjects factors is that the variance of these factors is difficult to estimate if it is not provided by the authors. Variance estimation for each factor in a design is critical in the determination of the size of effect present.

In factorial designs the between-subjects variance is the within-cell variance, which can be easily calculated from the summary statistics that are provided in research articles. However, the variance for a within-subjects design is the variance of the difference scores. If this variance is not provided in a research report, it can not be calculated straightforwardly. This difficulty in determining the variance in within-subjects designs stems from the differing composition of the F-ratio depending on which mean square estimates provide the appropriate error term for that particular ratio. The error

terms utilized in the computation of this ratio are often difficult to determine from the published information. This difficulty is in direct contrast to the construction of F-ratios in between-subjects designs in which the error term is always the same mean square for each F-ratio that is computed. Since the F-ratio is constructed in the same manner for each between-subjects factor, the estimation of variance from these designs can be carried out using the same set of techniques for each depending on how much information is provided. The estimates of variance yielded from the decomposition of summary statistics will be on a consistent scale across all variables in the between-subjects designs and across all studies that utilize this type design. This is not the case in the within-subjects designs where a specific set of techniques for variance decomposition are difficult to employ given that the construction of F-ratios is not consistent and is often difficult to determine.

The within-subjects studies share an additional problem in terms of cumulating effect sizes across studies, e.g. that of scale. If the variance can be identified for a within-subjects factor it will not be on the same scale as the variance of other within-subjects factors from the same design or with the variance of factors from any other study. Effect sizes based on variances that differ in scale from variable to variable have limited utility in a quantitative cumulation of research findings.

In the past, researchers encountering repeated measures in construction of meta-analyses have removed them from consideration (Hunter, 1984; Mongeau, 1984). However, deception detection research is not an area of study where the issue of within-subjects factors can be coped with by simply excluding those studies which employ repeated measures from cumulation of data. Questions of interest in deception detection studies are often directed at factors affecting the accuracy in veracity judgments. The designs of these studies typically match these research questions by asking the same observers to make such decisions across a number of different observation conditions, over a number of different deceivers/truthtellers or with differing amounts of information. This type of design results in judgments being made by the same individuals across different conditions, e.g. a design of repeated measures.

Extraction of Effect Sizes for Repeated Measures: The Special Case

Lacking complete summary statistics, estimation of effect sizes for within subjects factors is not possible. While this dissertation does not provide an answer to the problem of assessing within-subject effect sizes for all cases, it does present a solution to cumulation of within-subject effect sizes in a specific case. The special case addressed in this study is that of measurements of phenomena recorded in naturally occurring dichotomous units.

The research on accuracy in deception detection freely lends itself to natural units in its assessment of detection accuracy. For example, researchers often address questions such as the following: "How many times was the observer correct in her judgments of the deceptiveness of communicators?" or "What was the percentage or proportion of correct judgments made by the observer?" These questions and those like them can be easily measured with counts of correct judgments that can be converted into percents and proportions.

Conceptually, it is easy to see why this form of measurement is popular with deception researchers. Researchers who report their results in terms of counts, proportions or percentages derive them by asking observers to make the the dichotomous decision of "truth", if they believe the communicator is telling the truth and "lie", if they believe the communicator is lying to them. The observers' judgments then are compared to the actual behavior of the communicator: i.e. Did the communicator actually lie or tell the truth? The correct judgments are then tallied and the results presented as proportions or percentages of correct judgments or as a simple count of the number of times the observers were correct. Measurements of accuracy in deception detection reported in these natural units easily lend themselves to a straightforward method of meta-analysis that is appropriate for cumulation of study results for both within-

subjects and between-subjects variables. The repeated measures and mixed designs of deception detection research, along with the occasional between-subjects designs, can all be assessed using this method providing the accuracy scores are reported in percentages, proportions or counts. Results reported in counts can easily be converted by the reader to proportions by noting how many times the observers are asked to make judgments and comparing the number of judgments on which the observer responded correctly.

These naturally occurring measures provide the meta-analyst with a constant unit magnitude that is scaled both within studies and across studies. This constant unit can be used for cumulative comparison, a type of comparison currently not possible with estimates from within-subjects designs that do not use this form of measurement.

Use of proportions or percentages for cumulation focuses on the mean values for comparisons. This avoids drawing upon the variance of differences from within-subjects factors for comparison with the variance of differences from other within-subject factors or for comparison with the within-cell variances of between-subjects factors. Comparison of means with the same scale of magnitudes allows the researcher to avoid the problem of differing variance composition that are inherent in the comparison of effect sizes computed with non-comparable variances.

Proportions or percentages of correct veracity judgments measured on a dichotomous rating of truth or lie create a scale with possible values ranging from 0 to 1 with chance accuracy represented by .5 or 50 percent. One benefit of this type of scale, which may have drawn primary researchers to its use, is that accuracy in deception detection can be easily understood. A cumulative answer of .5 to the question of how accurate humans are at detecting deception can easily be translated as meaning people are correct in their judgments about half of the time. This percentage/proportion scale is equally as useful for understanding a question posed in a primary study as it is for understanding a question posed in a meta-analysis.

The use of mean percent accuracy by Kraut (1980) as an appropriate cumulative statistic in his quantitative summary was a good decision; however, the meta-analysis presented will avoid errors made by Kraut: poor study sample selection, inclusion of studies containing measurement scales and means that are not comparable, and failure to look at the differing impact of experimental conditions on the ability to detect deceptive communication.

Sample Selection

The studies included in this meta-analysis were extracted from extensive computer and manual searches through library indices, abstracts and through investigation of the supporting references that accompanied deception detection articles. The measurement scales of the located articles were then assessed in terms of the appropriateness for inclusion in this meta-analysis.

The studies chosen for this meta-analysis were selected because they included measurements of deception detection rated on dichotomous scales of truth or lie that were converted to proportions or percentages, or they included sufficient information to convert the counts into proportions or percentages.

The final requirement placed on the studies selected was that they contain a counterbalanced manipulation of truth and lies so that each person observed an equal number of truthful messages and deceitful messages. Specifically, the ability to detect deception can be defined as the ability to discriminate lies from the truth. This ability can be tested by having observers try to discriminate lies from truths when both are present. Asking an observer to discriminate lies from truths when only lies are present does not measure deception detection ability because the measurement is confounded with the variable of suspicion. In such a test of lie detection ability, highly

suspicious people will make the most judgments of deception and hence will have the highest accuracy rating. These people may actually have very poor lie detection ability as defined in this meta-analysis, as they may consistently make the error of suspecting truthful messages are deceptive. Experiments that only supply truthful messages also provide a confounded measure of deception detection. In this case very trusting or gullible people would have the highest accuracy scores. Again these people may be very poor at discriminating lies from the truth, trusting not only those who are telling the truth, but also those who are deceiving.

Thirty-two studies were found that conformed to these specifications (see Table 1) representing a total of 3,439 observers. This represents approximately two thirds of the observers examined in deception detection research.³ Studies that were not included in this meta-analysis were excluded for methodological reasons.

Excluded from this meta-analysis is lie detection research that relied on Likert-type scales ranging from 1 to 7 or 1 to 10. These studies, such as those by DePaulo, Rosenthal, Green and Rosenkrantz (in press), Geis and Moon (1981), Hemsley and Doob (1978), and Rotkin (1980) measured observer veracity judgments in terms of degree of certainty in their judgments. For example, a

Table 1
Studies Selected for Meta-Analysis

No	Year	Authors
1	1976	Atmiyanandana
2	1977	Bauchner, Brandt & Miller
3	1980a	Brandt, Miller & Hocking
4	1980b	Brandt, Miller & Hocking
5	1982	Brandt, Miller & Hocking
6	1974	Ekman & Friesen
7	1941	Fay & Middleton
8	1978	Harrison
9	1977	Hemsley
10	1979	Hemsley & Doob
11	1979	Hocking, Bauchner, Kaminski & Miller
12	1977	Lavrakas
13	1979	Littlepage & Pineault
14	1981	Littlepage & Pineault
15	1982	Littlepage & Pineault
16	1983	Littlepage, McKinnie & Pineault
17	1976	Maier & Lavrakas
18	1967	Maier & Janzen
19	1968	Maier & Thruber
20	1970	Matarazzo, Wiens, Jackson & Manaugh
21	1983	Miller, deTurck & Kalbfleisch
22	1974	Motley
23	1978	Parker
24	1982	Potamkin
25	1982	Rovira
26	1981	Sakai
27	1981	Sereno
28	1984	Stiff & Miller
29	1975	Wilson
30	1984	Zuckerman, Koestner & Alton
31	1984	Zuckerman, Kernis, Driver, & Koestner
32	Press	Zuckerman, Koestner, & Colella

rating of "1" represented extreme certainty that the communicator was lying, "2" indicated some certainty the communicator was lying, "3" a guess the communicator was lying, "4" uncertain whether the communicator was lying or telling the truth, "5" a guess the communicator was telling the truth, "6" somewhat certain the communicator was telling the truth, and "7" extremely certain the communicator was telling the truth. Measurements of lie detection accuracy rated on scales from (1) to (10) were similar with inclusion of more degrees of certainty in veracity judgments.

The basic assumption of researchers that measure lie detection accuracy thus is that lie detection is a matter of degree. Therefore, high accuracy scores can be achieved by extremely certain judgments that are correct. People who are extremely certain in their correct decisions receive higher accuracy scores than people who are somewhat certain and correspondingly people who are extremely certain in their incorrect decisions are less accurate than people who are less certain of their incorrect decisions. The weakness of using this measurement scale to assess accuracy in deception detection is that the observers accuracy judgments are confounded with an assessment of personal assuredness in their decision.

Research by Miller and Kalbfleisch (1982), Hocking (1976), and Littlepage and Pineault (1979) indicates that the

relationship between accuracy in veracity judgments and confidence in these judgments is weak. Hocking (1976) reported correlations between accuracy and confidence ranging from .061 to .063. Miller and Kalbfleisch (1982) noted that high confidence scores accompanied low accuracy ratings. Miller and Kalbfleisch also found that observers were more confident in their judgments of women than in their judgments of men.

Inclusion of a measure of confidence into an accuracy measure such as in the DePaulo et al. (in press), Geis and Moon (1981), Hemsley and Doob (1978) and Rotkin (1980) studies, results in an accuracy measure that is confounded by an unrelated variable. For example, this scale will measure cautious observers that make mostly accurate veracity decisions as less accurate than observers who are extremely certain on a few correct judgments and unable to make a judgement of truth or deceit on the remainder of the messages. On the other hand, some observers may express extreme confidence in their judgments of truth, but be cautious in their attributions of deceit. Researchers finding significant differences in observers' ability to detect deception in male as opposed to female communicators through use of likert type scales may actually be finding sex related differences in judgmental confidence and not differences in discerning truth from lies.

Studies that utilized a second technique to assess accuracy in deception detection were also excluded from this meta-

analysis. These studies (DePaulo, Davis & Lanier, 1980; DePaulo Lanier & Davis, 1983; DePaulo, Lassiter & Stone, 1982; DePaulo & Rosenthal, 1979; DePaulo, Stone & Lassiter, submitted; Olson, 1978; and Streeter, Krauss, Olson & Apple, 1977) had observers rate their veracity decisions on scales from 1 to 6 or from -3 to +3 with the lower rating indicative of extreme certainty the communicator was lying and the highest rating indicative of extreme certainty the communicator was telling the truth. The ratings between these extremes represented lesser degrees of certainty in the observers' decisions. Researchers using this method then subtract observers' ratings of truthfulness when the communicators were telling the truth from observers' ratings of truthfulness when the communicators were lying. The resulting value was used to represent the accuracy score of each observer. These accuracy scores are confounded by confidence in judgments as were the untransformed Likert measures of accuracy. Extreme confidence in a correct judgment is also assumed to be a more accurate assessment of veracity than a less confident correct judgement.

Some studies have tried to measure deception indirectly by measuring differences in pleasantness (Feldman, 1979; Feldman, Jenkins & Popoola, 1979), and differences in genuineness feedback (Feldman, 1979). For example, Feldman (1979) and Feldman et al. (1979) asked judges to rate the pleasantness expressed by

children as they drank Kool-Aid that was either sweetened (pleasant) or unsweetened (unpleasant). Children were asked to either express their actual satisfaction with a drink or to express their opposite reaction to it. Pleasantness was then rated on a six point scale by the judges and the experimenters converted these scores into difference values by subtracting ratings of pleasantness when the children were lying from ratings of pleasantness when the children were being honest in their expression of emotion. The inference in these studies is that ratings of pleasantness/unpleasantness can be converted into ratings of truth/lie so that accuracy in deception detection can be inferred.

There are several problems with this inference. First, children may differ in their abilities to communicate pleasantness and unpleasantness. Charlesworth and Krevtzer (1973) have noted that children are at lower levels of cognitive development and have less control over their facial muscles than adults. These children may not have been able to express the emotions the researchers expected of them. Second, the deception detection judges may differ in their definitions of pleasant and unpleasant behavior. This may have resulted in unreliable measurement. Third, the children may not have perceived the sweet drinks to be pleasant and the unsweetened drinks to be unpleasant. Hence, the researchers may have been unaware whether children were actually lying or telling the truth. Finally,

"unpleasantness" is one of the cues Mehrabian (1977) cites as being indicative of deceptive communication. Therefore, children being truthful about the unpleasantness of the drink may have been more likely to be judged as deceptive than those being deceptive about the unpleasantness of the drink. Conversely, children lying about the pleasantness of a drink may have been judged as more deceptive than children being truthful about the pleasant taste.

Studies using indirect measures of deception and truth such as those used in Feldman (1979) and Feldman et al. (1979), have been excluded from this meta-analysis.

Other studies not included in this meta-analysis have idiosyncratic components that make them unsuitable for cumulation. The deception detection studies by Geizer, Rarick and Soldow (1977) and Zuckerman, Amidon, Bishop and Pomrantz (1982) were excluded because they used a trichotomous measure of accuracy instead of a dichotomous measure. For these studies chance accuracy would be .333 rather than .50. Percentage accuracy means from these studies can not be accurately cumulated with the percentage accuracy means generated from dichotomous judgmental choices and ultimately tested against an overall chance accuracy of .50.

Kraut and Poe (1980) and Fugita, Hogrebe and Wexley (1980) were also be excluded from this meta-analysis because neither

study reported the experimental means. Finally, Hildreth (1953) and Littlepage and Pineault (1978) are excluded. While these researchers provided observers with a dichotomous truth/lie measure, observers were also given the option to avoid making a decision if they were not sure of their judgment. The difficulty in including these studies is that the mean percentage accuracy is based on only confident judgments, unlike the other studies in the meta-analysis which did not exclude the judgments made in uncertainty. It is possible that by excluding uncertain judgments Hildreth (1953) and Littlepage and Pineault (1978) measured only those observers with confidence in their deception detection ability. Conversely it is also possible that the accuracy ratings from these studies is yielded from judgments based on only those communicators who presented themselves in an obviously deceitful or honest manner.

The studies of deception detection included in this meta-analysis are those that assessed accuracy by dichotomous truth/lie measures without a no judgment option. These studies all reported mean accuracy ratings in proportions/percentages or their results were convertible to proportions/percentages.

Procedures for Cumulation

The first step of this meta-analysis consisted of searching for deception detection studies and selecting studies for cumulation that met the specifications outlined above. Second, the mean accuracy score for each level of research design was

extracted. These mean accuracy ratings were then placed into tables with other means taken from studies that measured these same variables. These tables are arranged according to the combinations of variables and levels that conceptually form general designs for the deception detection studies. Entries in these tables are weighted and averaged to provide cumulative estimates of the effect of each variable. Information is provided on the number of observers that each mean effect size represents. Means from deception detection studies that investigate variables not considered in other experiments are presented narratively after the cumulative estimates have been discussed.

In evaluating the effect sizes yielded from this meta-analysis, the size of each cumulative mean will be considered relative to the chance criterion of .5. Since variances can not be determined for the means represented in these cumulative estimates, significance tests of the difference between the means and .5 are not possible. Instead visual comparisons of relative size are made. In the analysis of cumulative means, these comparisons are made in light of the sizes of accuracy ratings common to this area of research. Since variance estimates are unavailable, sampling error can not be assessed in this meta-analysis.

RESULTS

The results of this meta-analysis indicate that humans are not very skilled at detecting deception. The cumulative accuracy ratings cluster from .45 to .70 with only a few cumulative accuracy ratings surpassing or falling behind these scores. These results also suggest that lie detection ability varies only slightly across experimental conditions. Keeping these modest differences in mind, the design tables from this meta-analysis will be evaluated for those variations present in the human ability to detect deception.

Observational Conditions

The first three tables of this meta-analysis examine the support for two major theories of deception detection: Ekman and Friesen (1969, 1974) and Maier and Thurber (1968).

Ekman and Friesen (1969, 1974) suggest that people are less aware of their bodies and extremities than they are their faces. They contend that since the face has a larger message sending capacity, most people learn to control their facial movements in order to accurately communicate with others. Consequently, when people are confronted with the task of concealing or distorting messages they will exercise control over their facial regions. Accordingly since the body and extremities have less sending

capacity for messages (Ekman & Friesen, 1969), people will have used them less to convey meaning and therefore be less aware of non-facial movement when they are concealing or distorting messages. Ekman and Friesen reason that given this tendency of communicators to concentrate on control of the facial regions and to remain unaware of other body movement, observers attempting to detect deception should focus on changes in body movement and to place less emphasis on facial displays.

Table 2 displays the cumulative results of studies that contained experimental conditions that test this theory. The observational condition containing full shots of both heads and bodies was added to the design to examine the impact of full visual information in comparison to the body only and head only conditions.

While the differences between these conditions appear small, the experimental condition of body only, where persons were allowed to observe only the bodies of communicators, yields the highest mean accuracy ratings. Conversely, observers were the least accurate in their observations of the head only condition. This cumulation of mean accuracy ratings supplies some support for Ekman and Friesen's theory. Specifically, observers in the viewing condition that Ekman and Friesen posit displays the most deception clues, the body, were more accurate than those observers who viewed the area Ekman and Friesen maintain is the most readily controlled, the face.

Table 2

Deception Detection Accuracy: Conditions Under Which Truthful
and Deceptive Messages are Observed⁴

No Study	Head Only	Body Only	Head & Body
2 Bauchner, Brandt & Miller			.47
6 Ekman & Friesen	.46	.52	
11 Hocking, Bauchner, Brandt & Miller	.51	.51	.52
13 Littlepage & Pineault	.53	.76	
14 Littlepage & Pineault	.49		
29 Wilson	.49	.61	.60
32 Zuckerman, Koestner & Colella	.59		
Weighted Mean	.51	.54	.53
No. of Observers Represented by Estimate	1395	1169	984

Those observers who viewed full head and body shots were more accurate than those observers that viewed heads only, but were less accurate than those observing bodies only. Deception clues supplied by the body, in the combined head and body condition, may enhance accuracy ratings when compared to ratings yielded from head only observations. On the other hand, full body and head information may inhibit accuracy ratings achievable when viewing bodies only. In this instance, the face, which is readily controlled by the communicator, may present contradictory information to the body's messages that confuses the observers

or facial cues may be attended to more readily by observers than cues from the body.

Maier and Thurber (1968) theorize that observers are overwhelmed by verbal, nonverbal, and content information when they decode messages. This overabundance of information distracts the observers from noticing clues that indicate they are being deceived. Maier and Thurber reason that deception clues will be more apparent if nonverbal information is reduced.

In Maier and Thurber's test of this theory, three experimental conditions were created: 1) watchers (who watched and listened to deceptive/truthful messages), 2) listeners (who listened to deceptive /truthful messages), and 3) readers (who read transcripts of deceptive/truthful messages). Their study found that those who listened and read were the most successful at detecting deception .77 in both cases. Observers who both watched and listened were less successful in deception detection with a mean accuracy rating of .58. Based on these results, Maier and Thurber reasoned that to accurately detect deception, communication should not be visually observed.

Table 3 presents the cumulative results of studies which test the experimental conditions of transcripts, audio only, and audio/visual. The results presented in Table 3 expand the conditions originally tested by Maier and Thurber by adding a

Table 3

Deception Detection Accuracy: Conditions Under Which Truthful
and Deceptive Messages are observed

No Study	Visual Only	Audio Only	Audio/ Visual	Tran- Script
2 Bauchner, Brandt & Miller	.47	.32		.47
6 Ekman & Friesen	.49			
7 Fay & Middleton		.56		
11 Hocking, Bauchner, Kaminski & Miller	.48	.54	.55	.57
13 Littlepage & Pineault			.64	
14 Littlepage & Pineault	.49	.63		
19 Maier & Thurber		.77	.58	.77
22 Motley		.67		
26 Sakai	.51	.58	.58	
29 Wilson	.62	.53	.57	
31 Zuckerman, Kernis, Driver & Koestner			.51	
32 Zuckerman, Koestner & Colella	.56	.62	.62	
Weighted Mean	.51	.58	.57	.61
No. of Observers Represented by Est.	1623	1676	1536	1018

condition of visual only, where the observers can see but not hear deceptive and truthful messages. This addition, adopted by some primary researchers, balances the audio only condition in this design.

The cumulative estimates reported in Table 3 indicate observers were the most accurate at detecting deception when they read transcripts of deceptive/truthful messages. These results support Maier and Thurber's contention that reduction of

nonverbal information will increase accuracy in detecting deception. In this condition both visual and paralinguistic cues were unavailable to the observers. The cumulative findings reported in Table 3, also suggests that persons with access to only audio cues and persons with access to both audio and visual cues have the same accuracy rates. This observation does not support Maier and Thurber's position that due to absence of distracting nonverbal information, observers in the audio only condition will be able to perform better than those in the full information condition. The new condition added to Maier and Thurber's original three, displayed the lowest accuracy ratings. When compared to the full audio visual condition, the low rating in the visual only condition suggests that the audio channel provides helpful information for the detection of deception.

Table 4 incorporates both the head and body conditions on which Ekman and Friesen concentrated their theory, with the audio, audio/visual and transcript conditions focused on by Maier and Thurber. The full head and body condition and the visual only condition added to Table 2 and Table 3 are also included.

This combined design suggests that persons are the most accurate in detecting deception when observing 1) the transcript only condition, 2) the audio only condition, 3) the body only condition with full audio information, 4) the combined head and

body viewing condition with full audio information, and 5) the head only condition with full audio information. The three conditions with the lowest accuracy ratings are: 1) the body only viewing condition without sound (6th), 2) the combined head and body viewing condition without sound (7th) and finally, 3) the head only viewing condition without sound (8th).

In general this table indicates that observers were more accurate in observational conditions that contained the content of the message than observational conditions that did not. These higher accuracy scores may have resulted from additional information available in message content. This information may have allowed observers to check messages for logic, consistency and length.

Observers viewing body only shots were more accurate than those viewing heads only. However, when observers viewed combined head and body shots, they were less accurate than when viewing bodies only and more accurate than when viewing heads only. While the body may be a source of helpful clues to deception as Ekman and Friesen contend (1969, 1974), this may only be the case when it is viewed alone without a view of the face and head.

Table 4

Accuracy in Deception Detection: Conditions Under Which Truthful and Deceitful Messages are Observed⁶

No	Study	Head Only		Body Only		Head & Body		Head & Body		Audio Only		Trans-Script
		Audio	No-Audio	Audio	No-Audio	Audio	No-Audio	Audio	No-Audio	Only	Only	
2	Bauchner, Brandt & Miller					.57	.47			.32		.47
6	Ekman & Friesen		.46		.52					.56		
7	Fay & Middleton									.54		.57
11	Hocking, Bauchner, Kaminski & Miller	.54	.47	.53	.48	.56	.49			.54		
13	Littlepage & Pineault	.53		.76						.63		
14	Littlepage & Pineault	.64	.49			.58				.77		.77
19	Maier & Thurber									.67		
22	Motley									.58		
26	Sakai	.58	.51							.53		
29	Wilson		.49		.61	.57	.62			.53		
32	Zuckerman, Koestner & Colella	.62	.56							.62		
Weighted Mean		.56	.48	.58	.51	.57	.50	.58				.61
No. of Observers Represented by Estimate		1157	1543	751	1137	1203	984	1676				1018

Familiarity

Some researchers have examined whether individuals are more accurate in deception detection when they have some knowledge regarding the typical communication style of the person they are judging (Brandt, Miller & Hocking, 1980a; Brandt, Miller & Hocking, 1980b; Brandt, Miller & Hocking, 1982; Ekman & Friesen, 1974). They reason that familiarity with this style should provide observers with a frame of reference from which deviations in communication patterns can be spotted.

In these experiments familiarity was operationalized by showing observers a short segment of communicators telling the truth. Observers were told that they were viewing truthful samples of behavior. These truthful segments were shown prior to the message segments where observers were asked to detect lies and truths.

Table 5 presents the cumulative results from studies that provided observers with truthful baseline messages. These truthful baselines were either not shown to observers, or they were shown one, two, three or six times prior asking subjects to make veracity assessments. This table reveals that in all conditions observer given a frame of reference were more accurate

Table 5
Baseline Familiarity With Communicator and Judges Accuracy in Deception Detection

No	Study	No Familiarity	Familiarity 1 Exposures	Familiarity 2 Exposures	Familiarity 3 Exposures	Familiarity 6 Exposures
3	Brandt, Miller & Hocking	.38	.51		.59	.44
4	Brandt, Miller & Hocking	.42			.66	
5	Brandt, Miller & Hocking	.31	.52	.56		
6	Ekman & Friesen	.47	.51			
Weighted Mean		.42	.51	.56	.64	.44
No. of Observers Represented by Estimate		482	382	99	150	50

in detecting deception than observers without this base for comparison. Observers who viewed the truthful baseline messages three times were the most accurate, followed by those who viewed the baseline two times, and those who viewed the baseline once. This pattern seems to indicate that the more familiar observers are with a communicator's idiosyncratic truthful behavior, the better able they are to detect deviations from this behavior, and in turn, the better they are able to detect deceptive communication.

While this pattern of increased accuracy ratings is evident for baseline observations of up to three message repetitions, observers who viewed truthful baseline messages six times had lower accuracy ratings than any of the other conditions where the familiarity segment was provided. An explanation for this decreased accuracy may be that observers in this condition became fatigued or bored watching the same segment six times. This boredom or fatigue may have reduced observers efficiency in information processing.

Sex Differences in Deception Detection

Male versus female ability to detect deception has also been explored by prior researchers. Table 6 displays the results of a cumulation of studies that examined sex differences in detecting deceit. The cumulative mean accuracy ratings suggest that

Table 6
Judges Sex and Accuracy in Deception Detection

No	Study	Male	Female
1	Atmiyanandana	.55	.52
7	Fay & Middleton	.55	.56
9	Hemsley	.51	.54
19	Maier & Thurber	.69	.74
23	Parker	.51	.50
25	Rovira	.52	.54
26	Sakai	.54	.57
Weighted Mean		.59	.61
No. of Observers Represented by Estimate		331	327

females are better at detecting deception than males. The results are consistent with nonverbal research on sex differences and decoding ability. Hall (1978,1980) has noted that females are more skilled in decoding nonverbal messages than men. This heightened sensitivity to nonverbal cues that allows females to more accurately interpret their meaning may also be of use in detecting discrepant nonverbal cues or other such clues that deception may be occurring. Tables 7 and 8 break down this detecting ability across observation conditions for males and females respectively.

Table 7

Deception Detection Accuracy for Male Judges

No	Study	Visual	Audio	Audio/ Visual	Tran- Scripts
19	Maier & Thurber	.57	.76		.73
26	Sakai	.49	.57	.57	
Weighted Mean		.54	.68	.57	.73
No.of Observers Represented by Est.		205	205	90	115

Table 8

Deception Detection Accuracy for Female Judges

No	Study	Visual	Audio	Audio/ Visual	Tran- Scripts
19	Maier & Thurber	.63	.78		.80
26	Sakai	.53	.60	.59	
Weighted Mean		.58	.70	.59	.80
No.of Observers Represented by Est.		194	194	90	104

The observation conditions examined in these tables appear to have similar moderating effects on the accuracy of both male and female lie-detectors. However, in each observation condition female observers maintained higher rates of success in detecting deception.

Sex differences in deception detection ability have also been studied in terms of the sex of the communicators for whom male and female observers make judgments. The rationale for these comparisons in primary research centered on whether observers were better able to identify lies and truths perpetrated by same or opposite sex communicators (i.e. Parker, 1978; Rovira, 1982). Results cumulated from these studies are presented in Table 9 and Table 10.

Table 9

Male Judges: Deception Detection Accuracy Observing
Male vs Female Communicators

No	Study	Male	Female
7	Fay & Middleton	.57	.53
9	Hemsley	.51	.51
23	Parker	.46	.49
25	Rovira	.47	.56
Weighted Mean		.50	.52
No. of Observers Represented by Estimate		90	90

Table 10

Female Judges: Deception Detection Accuracy Observing
Male vs Female Communicators

No	Study	Male	Female
7	Fay & Middleton	.58	.54
9	Hemsley	.51	.56
23	Parker	.46	.52
25	Rovira	.50	.58
Weighted Mean		.50	.54
No. of Observers Represented by Estimate		97	97

The cumulative means show that both male and female judges are more accurate when they are judging females than when they are judging males. While female observers were better able to judge members of their own sex and males were better able to judge members of the opposite sex this finding may be more related to the deceptive and truthful communication patterns of females and males, and less related to specific observer skills in detecting members of their own or opposite sex.

Table 11
Sex of Communicators and Accuracy in Deception Detection

No	Study	Male	Female
7	Fay & Middleton	.58	.54
9	Hemsley	.51	.53
12	Lavrakas	.52	.58
23	Parker	.47	.50
25	Rovira	.49	.57
Weighted Mean		.51	.54
No. of Observers Represented by Estimate		287	287

Table 11 presents the combined ratings of male and female observers broken down by the sex of the communicators who were being judged. These results show that observers may be able to more accurately detect deception by females than by males.

Judgments of Truth vs. Lie

The studies included in this meta-analysis were all counterbalanced so that observers were presented with equal numbers of truths and lies. Only counterbalanced studies were included in order to reduce the effect of guessing bias on the accuracy measures. Table 12 shows this preference for judgments of truth or deceit. From this cumulative finding it appears that observers are more likely to judge communicators to be telling the truth than to judge them to be lying.

Table 12
Observer Judgments of Truth vs Lie

No	Study	Truth	Lie
6	Ekman & Friesen	.45	.53
7	Fay & Middleton	.50	.61
8	Harrison	.75	.48
12	Lavrakas	.50	.60
13	Littlepage & Pineault	.68	.60
15	Littlepage & Pineault	.81	.30
16	Littlepage, McKinnie & Pineault	.57	.60
18	Maier & Janzen	.34	.50
19	Maier & Thurber	.71	.71
24	Potamkin	.68	.58
28	Stiff & Miller	.67	.41
31	Zuckerman, Kernis, Driver & Koestner	.60	.41
32	Zuckerman, Koestner, & Colella	.60	.58
Weighted Mean		.59	.55
No. of Observers Represented by Estimate		1233	1233

Narrative Review

The studies included in this narrative review met the sampling requirements of the meta-analysis for measurement and counterbalancing. Even so these studies can not be cumulated because few other studies examining the same variables are available, or meet the sampling requirements. The low accuracy ratings in the cumulative results are also prevalent in these

findings. The differences between experimental conditions are also small. These estimates are based on smaller samples than most of the cumulative findings and the variations suggested in this narrative review should be interpreted with caution.

Deceiver differences: rehearsal, self-monitoring age and drugs. Researchers have explored the detectability of communicators who have rehearsed their truthful and deceptive messages (Littlepage & Pineault, 1982; Miller, deTurck & Kalbfleisch, 1984). In their 1982 study, Littlepage and Pineault found that if observers were given time to plan their messages, observers had more difficulty detecting deception (.52) than when they were not given rehearsal time (.59).

Miller et al. (1983) further explored the impact of planning on observer success in deception detection by adding a self-monitoring measure (Snyder, 1974) to this design. In this study, message rehearsal time differentially affected detectability depending on the degree to which communicators self-monitored. High self-monitors were more difficult to detect when given time to plan their messages (.45) than when not given time to rehearse (.50). Conversely, low self-monitors were easier to detect after given time to rehearse (.56) and more difficult to detect when their lies were spontaneous (.53). These researchers reasoned that high self monitors had greater confidence in their performance abilities, and therefore used the planning time to

rehearse and better prepare their responses to an interviewer's questions. On the other hand, low self-monitors were not confident in their performance abilities, and became nervous and apprehensive when given time to think about the messages they would soon have to communicate. Consequently, low self-monitors were not as successful as they would have been without the rehearsal time. In both the rehearsal and no rehearsal conditions high self-monitors were more difficult to detect than low self-monitors.

Age may also affect the detectability of deceivers. Parker (1978) found the combination of sex and age differentially affected observer success in deception detection. Specifically, thirteen to fourteen year-old females were easier to detect (.52) than adult females over eighteen (.50) and seven-year-old female children (.48). Conversely, thirteen to fourteen year old males were easier to detect (.48) than seven-year-old males (.47) and adult males over eighteen (.45).

Potamkin (1982) studied deception by heroin addicts. Given the experience addicts have had in hiding their habit from others, she reasoned that addicts should be harder to detect than nonaddicts. However, the study found that addicts were easier to detect (.64) than were nonaddicts (.62). It appears from this study that despite their experience in deceiving others about their habits, addicts are no more skilled than others.

Message characteristics. The type of message also appears to affect how accurate observers will be in deception detection. Hocking, Bauchner, Kaminski and Miller (1979) found that observers more accurately detected factual lies (.54) than lies about emotional states (.50). Rovira (1982) found that observer accuracy was affected by whether or not they agreed with the content of a communicator's message. The results of Rovira's study suggest that observers are more accurate in detecting deception when they agree with the message (.53), than when they disagree (.52). Finally, Maier and Lavrakas (1976) found that persons observing messages that a polygraph had correctly identified, were better able to detect deception (.68), than when they observed messages which the polygraph was unable to correctly identify (.51).

Social-cultural observer differences. While this meta-analysis has suggested that females are slightly better at deception detection than males, the primary research by Parker (1978) suggests that this ability may differ according to age. In this study adult females over eighteen were better at lie detection (.51) than seven-year-old females (.50), and thirteen to fourteen year old females (.49). Thirteen to fourteen year old males, on the other hand, were more accurate (.52) than were seven-year-old males (.50) and adult males over eighteen (.50). According to the results yielded from this study, females are somewhat better lie detectors than males when they are adult,

equal in lie detection ability as children and worse than males as teenagers.

Instead of examining observer age, Sereno (1981) looked at past experience in his study of success in detecting the truths and lies of children. This study found that elementary school teachers were the most accurate in judging children (.77). Sereno had posited that those observers with the most experience with children would be the best at determining when they were lying and telling the truth. While teachers had the highest accuracy rates, adults who were parents of children were the least accurate in detecting the lies of children (.63). Adults with no children were slightly more accurate than the adults with children (.64). In interpreting these results, it could be reasoned that teachers were better able to detect children's deception because of the wide range of children they must work with on a daily basis. This experience may have made them aware of children's typical styles of communicating, which in turn could have allowed them to better discern when the children were deviating from these patterns. Conversely, the parents of children may only have been exposed to the idiosyncratic communication behaviors of their own children, hence they could not generalize this knowledge as well as the teachers. The childless adults may also have been generalizing from limited knowledge.

In her study of heroin addicts discussed in the proceeding section, Potamkin (1982) examined the differential deception detection abilities of addicts and nonaddicts. Potamkin found that nonaddicts were slightly more accurate at detecting deception (.64) than were addicts (.62).

Finally, Atmiyanandana (1976) considered differences in deception detection accuracy by observers from different nationalities. In his study Asian, Latin American, and North American observers were assessed. Atmiyanandana found that Asian judges were the least successful with an accuracy rate of .50, while both North and Latin Americans had a slightly higher accuracy rate of .54.

Experimentally induced observer differences. Several researchers have studied the impact of various experimental manipulations on ability to detect deception. For example, Motley (1974), explored whether telling observers to attend to a specific nonverbal cue would increase their accuracy in deception detection. In his study Motley told half his observers to pay particular interest to speech latency while listening to audio tapes. Results indicated that the observers told to attend to speech latency were less accurate (.63) than those who were given no listening clues (.67). This study could be interpreted as indicating that directing observers' attention to specific cues associated with deception does not increase the ability to spot deceivers and truth-tellers. However, it may also be the case

that speech latency by itself is not that helpful an indicator of deception, and other cues more strongly associated with deception might produce greater observer accuracy.

Zuckerman, Koestner and Alton (1984) attempted to teach observers how to spot lies and truth by showing them audio/visual tapes of communicators and telling them when the communicators were lying and when they were telling the truth. In this study, the accuracy of observers given no instructions was .62, while the accuracy ratings for the teaching conditions ranged from .61 to .70 depending on the instruction method used.

The highest accuracy rating in Zuckerman et al. (1984) was shared by observers in two experimental conditions. In one of these conditions experimenters showed observers tapes of eight communicators. After each tape the researchers had observers make veracity judgments followed by instructions regarding which communicators were lying or telling the truth. These observers reached an accuracy level of .70, with accuracy computed on all eight judgments. The other condition that achieved a .70 accuracy rating was one in which observers were told before the first four tapes which communicators were telling the truth or lying, and told after the last four tapes which communicators were lying or telling the truth. Accuracy scores were computed on judgments of the last four tapes. Observers told which communicators were lying or telling the truth before the first

four tapes but given no information for the last four tapes were accurate at the .66 level, with accuracy computed on the last four judgments. The results of these conditions suggest that observers may have been learning how to spot deceptive and truthful communication.

Observers in two conditions in the Zuckerman et al. (1984) study had accuracy levels below the control group. In one of these conditions observers were informed after the first four tapes which communicators were deceptive or truthful, but not after the last four tapes. Observers in this group were accurate at the .61 level, with accuracy computed on all eight judgments. In the experimental condition with the lowest accuracy, observers were given half truthful and half false feedback concerning which communicators were lying or telling the truth. This feedback was provided after each judgment was made on the eight communicators. Accuracy in this condition was .53, computed on all judgments. The Zuckerman et al. (1984) rationale for inclusion of this condition was to examine whether observers were actually learning how to spot deceivers, or whether they were simply making their judgments based on anticipated proportions of truth and deception. The same order sequence of truth and lie feedback was provided in this condition as was provided in the eight-after condition (with truthful feedback), which yielded the high accuracy score of .70. Zuckerman et al. reasoned that if observers were using feedback to calculate probabilities of

correct responses in both the veridical eight after group and in the false-eight after group their accuracy scores should be similar. But if observers were using the feedback to learn to spot truth and lies the observers with the veridical information should have superior accuracy scores to those in the false feedback condition. From the comparison of the false condition to the veridical one it appears that some learning may have been taking place among the deception judges.

Stiff and Miller (1984) did not provide observer with feedback, but they did provide observers with additional information not typically available in traditional deception studies. These researchers showed observers tapes of communicators lying and telling the truth in response to an interviewer's question. After the communicators had responded to this question the interviewer then further probed their response. These probes either suggested the interviewer believed the communicators or disbelieved them. The video tapes shown to the observers included the communicators' responses to these probes. Findings indicate that observers viewing communicators experiencing a positive probe (i.e. indicating the interviewer believed communicator) were able to detect detection at the same accuracy rate (.54) as observers who viewed communicators experiencing a negative probe (i.e. indicating the interview disbelieved subject).

Finally, most deception research has observers make judgments of the veracity of communicators from a vantage point in which they do not have to interact with the communicators. Instead of this popular observation position, Hemsley and Doob (1979), had 27 persons interview communicators and make veracity judgments. The interviewers were told to make their veracity decisions based on observation of behavioral cues, and not by attempting to trap the respondent with questions. These interviewers detected deception at a .58 level of accuracy.

Hemsley and Doob (1977) did not have a comparison condition for subjects observing communicators only. However, Matarazzo et al. (1970) examined this issue by having one person interview communicators and another person watch the interviews. In this study, the interviewer had an accuracy rate of .58 and the observer had an accuracy rate of .60.

In summary, the research concerning the experimental conditions discussed in this section is mixed. Directing observers' attention to deception clues was found to decrease accuracy. While providing observers with some training may increase accuracy. In general, positive and negative probes by interviewers did not differentially affect deception detection accuracy. Those interviewing deceivers also appeared to be less accurate than those who only observed.

DISCUSSION

Comparison with Previous Meta-Analyses

The findings of this meta-analysis are both similar and different from the findings generated by Zuckerman et al. (1981) and Kraut (1980). One of the major findings of this quantitative review was that humans were the most accurate in detecting deception when they had access to message content. Transcripts were found to yield higher accuracy than audio information, and the transcript only condition facilitated the highest accuracy ratings.

Zuckerman et al. also found that people were more accurate when they had access to message content. However these researchers found that observers exposed to audio information were more accurate in detecting deception than those who were exposed to transcripts only. Expressed in standard deviation units, accuracy scores yielded by the audio only condition had a mean of 1.09, and accuracy scores for the transcript only condition had a mean of .70. Zuckerman et al. also found that all visual observational conditions when combined with audio information yielded higher accuracy scores than the transcript only condition. Observations of the body only with audio had a mean of 1.49. Observations of the body and face with audio had a

mean of 1.00. And, observations of the face only with audio yielded a mean of .99. These accuracy scores, expressed in standard deviation units, suggest that the audio channel may be more useful for detecting cues to deception than the transcripts. Zuckerman et al. found all five of these observational conditions yielded accuracy scores significantly greater than chance.

A second major finding of this meta-analysis was that in the context of visual observations, shots of the body yielded the highest accuracy ratings. This condition was followed by accuracy ratings from the combined body and face condition, and the head only condition. This pattern was the same for visual conditions with audio and the visual conditions without audio; with visual conditions with audio having higher accuracy ratings than visual conditions without audio.

Zuckerman et al. also found this pattern. With audio information, body shots yielded the highest accuracy (1.49), followed by combined body and head shots (1.00) and shots of the head only (.99). Without audio information, body shots yielded the highest accuracy (.43), followed by combined body and head shots (.35), and shots of the head only (.05). These accuracy ratings, expressed in standard deviation units, were all found to be significantly greater than chance, except for the shots of the head only without audio.

Zuckerman et al. concluded their meta-analysis by indicating that observational conditions substantially moderate deception detection. These researchers further indicate that humans are more accurate than chance in detecting deception in all of the observation conditions with the exception of the head only condition without audio.

These statements seem to conflict with the results of this meta-analysis, which found human accuracy to be low and to not differ greatly from the chance rate of .50. The "substantial" differences in observational conditions are also in conflict with the slight variations found in this meta-analysis.

One explanation for these differences may be that the d statistic as operationalized by Zuckerman et al., is inflated and has provided estimates of accuracy much larger than the actual population values. This inflation might account for substantial differences and for the significant accuracy rates yielded in the Zuckerman et al. meta-analysis. Furthermore, these inflated values might also explain the differences in accuracy patterns for the transcript condition and audio conditions.

Other results from this meta-analysis and the Zuckerman et al meta-analysis are more similar. Both meta-analyses found that: 1) females were better able to detect deception than males, 2) observers could more easily detect female deceivers than male deceivers, and 3) high self-monitors were more difficult to detect than low self-monitors. This meta-analysis found these

differences to be slight, and Zuckerman et al. also found that the differences not to be significantly more than chance. Perhaps these differences were actually so small that even inflated d values could not make them appear substantial.

This meta-analysis found that familiarity with a truthful baseline of the person to be judged facilitates increased accuracy scores. The exception to this pattern was found when the baseline was repeated six times. In their narrative review, Zuckerman et al. also noted the same pattern. However, these researchers did not provide cumulative estimates of this relationship.

The meta-analysis presented here found that humans were not very accurate in detecting deception. While this finding is in contrast to the evaluation of deception detection by Zuckerman et al., it is similar to that of Kraut (1980). This meta-analysis shows that in general, Kraut's sole estimate of accuracy is correct in its low rating of the human ability to detect deception. However, this meta-analysis also suggests this rating may differ across experimental conditions and that important patterns in deception detection may be overlooked through the use of a singular estimate of human ability. For example, the facilitating impact of message content in accurately detecting deception can not be retrieved from the Kraut estimate. Neither

can the importance of the body, as a source of cues to deception be discovered in a single estimate.

Sampling Error

The results yielded by this meta-analysis have some uncertainty associated with them. This uncertainty is attributable to the sampling error present in the estimates. Actual accuracy rates of humans in detecting deception may differ somewhat from these estimates. Variation in deception detection ability across the experimental conditions may also be less than the present findings suggest.

Without the availability of full summary statistics for the studies included in this meta-analysis, the amount of sampling error present in the meta-analytic results can not be assessed. Meta-analysts have yet to determine how to assess sampling error with limited summary statistics for the within-subjects research designs.

While the amount of sampling error present can not be assessed, there are some indications that it may be less extensive in some of the estimates than in others, although the amount of this difference can not be ascertained. For example, those estimates with large cumulative sample sizes will have less sampling error than those with smaller samples. In this meta-analysis, the accuracy ratings for the observational conditions and for the judgments of truth and deceit will have the least

sampling error. The accuracy ratings with the most sampling error will be those yielded by the singular studies in the narrative review, and by cumulative ratings based on small sample sizes.

The fact that most of the studies comprising this meta-analysis have come from within-subjects designs suggests that the amount of sampling error present in these accuracy ratings is less than what it would have been had these studies come from factorial designs (Hunter, 1984). Sampling error is reduced in the repeated measures design, because individual differences in observer responses are not a source of error as they are in factorial designs. Instead these individual differences become a measure of the effect of the independent variables in the within-subjects design.

Conclusion

This meta-analysis has implications for both researchers engaged in the study of human deception detection, and in the development of improved research methodology.

For the deception researcher, this meta-analysis may provide some guidelines for future deception research. The most powerful guideline should come from the observational conditions. These conditions yielded the most distinct pattern in accuracy ratings and, with their large sample sizes, they

should contain the least amount of sampling error. Perhaps future deception research can further explore the viability of message content in facilitating increased deception detection accuracy. Second, the guessing bias found in judgments of truth and deceit should caution researchers to counterbalance truthful and deceptive messages. Finally, the low accuracy ratings evident in this meta-analysis may direct investigation into understanding how humans might best cope with poor lie detection skills.

For the research methodologist, this meta-analysis illustrates the need for methods of extracting effect sizes from within-subjects designs when complete summary statistics are unavailable. While an initial step in extracting effect sizes from within-subjects designs, this study provides only a partial solution to the problem. Techniques need to be developed for extracting effect sizes from studies that use a variety of measurement schemes, and which may provide differing amounts of summary information. Development of appropriate estimates for sampling error should also be considered.

References

- Atmiyanandana, V. (1976). An experimental study of the detection in cross-cultural communication. Doctoral dissertation, Florida State University. (University Microfilms No.76-29,416)
- Backster, C. (1963). Total chart minutes concept. Law and Order, 11, 77-79.
- Barland, G., & Raskin, D.C. (1976). Validity and reliability of polygraph examinations of criminal suspects (Report No. 76-1, Contract 75-NI-99-0001). Washington, DC: U.S. Department of Justice.
- Bauchner, J.E., Brandt, D.R., & Miller, G.R. (1977). The truth/deception attribution: Effects of varying levels of information availability. In B.D. Rubin (Ed) Communication Yearbook I, New Brunswick, New Jersey: Transaction Books.
- Ben-Shakhar, G., Liebllich, I., & Bar-Hillel, M. (1982). An evaluation of polygrapher's judgments: A review from a decision theoretic perspective. Journal of Applied Psychology, 67(6), 701-713.
- Bersh, P.J. (1969). A validation study of examiner judgment. Journal of Applied Psychology, 53, 399-403.
- Benusi, V. (1975). The response systems of lying. Polygraph, 4, 52-76. (Original work published 1914).

- Brandt, D.R., Miller, G.R., & Hocking, J.E. (1980a). The truth-deception attribution: Effects of familiarity on the ability of observers to detect deception. Human Communication Research, 6(2), 99-110.
- Brandt, D.R., Miller, G.R., & Hocking, J.E. (1980b). Effects of self-monitoring and familiarity on deception detection. Communication Quarterly, 28(3), 3-10.
- Brandt, D.R., Miller, G.R., & Hocking, J.E. (1982). Familiarity and lie detection: A replication and extension. The Western Journal of Speech Communication, 46, 276-290.
- Buck, R., Miller, R.E., & Caul, W.F. (1974). Sex, personality, and physiological variables in the communication of affect via facial expression. Journal of Personality and Social Psychology, 30(4), 587-596.
- Bugental, D.E., Kaswan, J.W., & Love, L.R. (1970). Perception of contradictory meanings conveyed by verbal and nonverbal channels. Journal of Personality and Social Psychology, 16(4), 647-655.
- Burgoon, J. & Saine, T. (1978). The unspoken dialog: An introduction to nonverbal communication. Boston: Houghton Mifflin.
- Charlesworth, W.W., & Krevtzer, M.A. (1973). Facial expressions of infants and children. In P. Ekman (Ed), Darwin and Facial Expression: A Century of Research in Review. New York: Academic Press.

- Cohen, J. (1977). Statistical power analysis for the behavioral sciences (Rev. ed.) New York: Academic Press.
- Comadena, M.E. (1981). Examinations of the deception attribution process of friends and intimates. Unpublished Doctoral Dissertation, Purdue University, Indiana.
- Comadena, M.T. (1982). Accuracy in detecting deception: Intimate and friendship relationships. In M. Burgoon (Ed.), Communication Yearbook 6. Beverly Hills: Sage.
- DePaulo, B.M. (1981). Success at detecting deception: Liability or skill? Annals New York Academy of Sciences, 245-255.
- DePaulo, B.M., Davis, T., & Lanier, K. (1980, April). Planning lies: The effects of spontaneity and arousal on success at deception. Paper presented at the Eastern Psychological Association, Hartford, Conn.
- DePaulo, B.M., & Jordan, A. (1982). Age changes in deceiving and detecting deceit. In R.S. Feldman (Ed.), Development of Nonverbal Behavior in Children, New York: Springer-Verlag.
- DePaulo, B.M., Jordan, A., Irvine, A., & Laser P.S. (1982). Age changes in the detection of deception. Child Development, 53, 701-709.
- DePaulo, B.M., Lanier, K., & Davis, T. (1983). Detecting the deceit of the motivated liar. Journal of Personality and Social Psychology, 45(5), 1096-1103.

- DePaulo, B.M., Lassiter, G.D., & Stone, J.I. (1982). Attentional determinants of success at detecting deception and truth. Personality and Social Psychology Bulletin, 8(2), 273-279.
- DePaulo, B.M., Lassiter, G.D. & Stone, J.I. (1982). Attentional determinants of success at detecting deception and truth. Personality and Social Psychology Bulletin, 8, 273-279.
- DePaulo, B.M., & Rosenthal, R. Telling lies. (1979). Journal of Personality and Social Psychology, 17(10), 1713-1722.
- DePaulo, B.M., & Rosenthal, R. (1979). Ambivalence, discrepancy, and deception in nonverbal communication. In R. Rosenthal (Ed.) Skill in Nonverbal Communication, Cambridge, Mass: Oelgeschlager.
- DePaulo, B.M., Rosenthal, R., Green, C.R., & Rosenkrantz, J. (in press). Diagnosing deceptive and mixed messages from verbal and nonverbal cues. Journal of Experimental Social Psychology.
- DePaulo, B.M., Rosenthal, R., Rosenkrantz, J., & Green C.R. Actual and perceived cues to deception: A closer look at speech. (Unpublished Manuscript)
- DePaulo, B.M., Stone, J.I., & Lassiter, G.D. (1984). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal deception success. Manuscript submitted for publication.

- DePaulo, B. M., Stone, J. I. & Lassiter, G. D. (in press).
Deceiving and detecting deceit. In B. R. Schlenker (Ed.)
The Self and Social Life. New York: McGraw-Hall.
- Dollinger, S.J., Reader, M.J., Marnett, J.P., & Tylanda, B.
(1983). Psychological-mindedness, psychological-construing,
and the judgment of deception. The Journal of General
Psychology, 108, 183-191.
- Edelman, R.I. (1970). Some variables affecting suspicion.
Journal of Personality and Social Psychology, 15(4), 333-337.
- Ekman, P., & Friesen, W.V. (1969). Non-verbal leakage and clues
to deception, Psychiatry, 32, 88-106.
- Ekman, P., & Friesen, W.V. (1974). Detecting deception from the
body or face. Journal of Personality and Social Psychology,
29(3), 288-298.
- Exline, R.V., Thibaut, J., Hickey, C.B., & Gumpert, P. (1970).
Visual interaction in relation to machiavellianism and an
unethical act. In R. Christie and F.L. Geis (Eds). Studies in
Machiavellianism, New York: Academic Press.
- Fay, P.J., & Middleton, W.C. (1941). The ability to judge truth-
telling, or lying, from the voice as transmitted over a public
address system. The Journal of General Psychology, 24, 211-
215.

- Feldman, M., & Thayer, S. (1980). A comparison of three measures of nonverbal decoding ability. The Journal of Social Psychology, 112, 91-97.
- Feldman, R.S. (1976). Nonverbal disclosure of teacher deception and interpersonal affect. Journal of Educational Psychology, 68(6), 807-816.
- Feldman, R.S. (1979). Nonverbal disclosure of deception in urban Koreans. Journal of Cross-Cultural Psychology, 10(1), 73-83.
- Feldman, R.S., Jenkins, L., & Popoola, O. (1979). Detection of deception in adults and children via facial expressions. Child Development, 50, 350-355.
- Feldman, R.S., & White, J.B. (1980). Detecting deception in children. Journal of Communication, 30(2), 121-128.
- Fugita, S.S., Hogrebe, M.C., & Wexley, K.N. (1980). Perceptions of deception: Perceived expertise in detecting deception, successfulness of deception and nonverbal cues. Personality and Social Psychology Bulletin, 6(4), 637-643.
- Geis, F.L., & Moon, T.H. (1981). Machiavellianism and deception. Journal of Personality and Social Psychology, 41(4), 766-775.
- Geizer, R.S., Rarick, D.L., & Soldow, G.F. (1977). Deception and judgment accuracy: A study in person perception. Personality and Social Psychology Bulletin, 3, 446-449.
- Glass, G.V., McGaw, B., & Smith, M.L. (1981). Meta-Analysis in Social Research, Beverly Hills: Sage.

- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. Psychological Bulletin, 85, 845-857.
- Hall, J. A. (1980). Gender differences in nonverbal communication skills. In R. Rosenthal (Ed.) Quantitative Assessment of Research Domains. San Francisco, California: Jossey-Bass.
- Harrison, A.A., Hwalek, M., Raney, D.F., & Fritz, J.G. (1978). Cues to deception in an interview situation. Social Psychology, 41(2), 156-161.
- Hemsley, G.D. (1977). Experimental studies in the behavioral indicants of deception. Unpublished Doctoral Dissertation, University of Toronto.
- Hemsley, G.D., & Doob, A.N. (1979). The detection of deception from nonverbal behaviors. Paper presented at the meeting of the Canadian Psychological Association, Quebec City, Quebec, Canada.
- Hemsley, G.D., & Doob, A.N. (1978). The effect of looking behavior on perceptions of a communicator's credibility. Journal of Applied Social Psychology, 8(2), 136-144.
- Hildreth, R.A. (1953). An experimental study of audiences' ability to distinguish between sincere and insincere speeches. Unpublished doctoral dissertation, University of Southern California.

- Hocking, J.E. (1976). Detecting deceptive communication from verbal, visual and paralinguistic cues: An exploratory experiment. Unpublished Doctoral Dissertation, Michigan State University.
- Hocking, J.E., Bauchner, J., Kaminski, E.P., & Miller, G.R. (1979). Detecting deceptive communication from verbal, visual, and paralinguistic cues. Human Communication Research, 6(1), 33-46.
- Horvath, F.S. (1977). The effects of selected variables on the interpretation of polygraph records. Journal of Applied Psychology, 62, 127-136.
- Horvath, F.S., & Reid, J.E. (1971). The reliability of polygraph examiner diagnosis of truth and deception. Journal of Criminal Law, Criminology and Police Science, 62, 276-281.
- Hunter, F.L., & Ash, P. (1973). The accuracy and consistency of polygraph examiners' diagnoses. Journal of Police Science and Administration, 1, 370-375.
- Hunter, J. E. (1982, January). A new design for psychological statistics. Unpublished manuscript, Michigan State University.
- Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). Meta-Analysis: Cumulating research findings across studies. Beverly Hills: Sage.
- Hunter, J.E., (1984, March). Personal communication.

- Jones, H.E. (1960). The longitudinal method in the study of personality. In I. Iscoe & H. W. Stevenson (Eds.), Personality development in Children. Chicago, Ill.: University of Chicago Press.
- Kepple, G. (1982). Design and analysis: A researchers handbook (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kleinmuntz, B., & Szucko, J.J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. American Psychologist, 39(7), 766-776.
- Kleinmuntz, B., & Szucko, J.J. (1982a) Is the lie detector valid? Criminal Defense, 9, 13-15.
- Kleinmuntz, B., & Szucko, J.J. (1982b). On the fallibility of detection. Law and Society, 17, 85-104.
- Kleinmuntz, B., & Szucko, J.J. (1984). A field study of the fallibility of polygraphic lie detection. Nature, 308, 449-450.
- Knapp, M.L., Hart, R. P., & Dennis, H. S. (1974). An exploration of deception as a communication construct. Human Communication Research, 1(1), 15-29.
- Kraut, R.E. (1978). Verbal and nonverbal cues in the perception of lying. Journal of Personality and Social Psychology, 36(4), 380-391.
- Kraut, R. (1980). Humans as lie detectors: Some second thoughts. Journal of Communication, 30, 209-216.

- Kraut, R., & Poe, D. (1980). Behavioral roots of person perception: The deception judgments of customs inspectors and laymen. Journal of Personality and Social Psychology, 39(5), 784-798.
- Kraut, R., & Lewis, S.H. (1984). Some functions of feedback in conversation. In H. Applegate and J. Sypher (Eds). Understanding Interpersonal Communication. Sage.
- Larson, J.A. (1932) Lying and its detection. Chicago: University of Chicago Press.
- Lavrakas, P.J. (1977). Human differences in the ability to differentiate spoken lies from spoken truths. Unpublished Doctoral Dissertation, Loyola University of Chicago.
- Lavrakas, P.J., & Maier, R.A. (1979). Differences in human ability to judge veracity from audio medium. Journal of Research in Personality, 13, 139-153.
- Levine, J.M., Romashko, T., & Fleishman, E.A. (1973). Evaluation of an abilities classification system for integration and generalizing research findings: An application to vigilance tasks. Journal of Applied Psychology, 58, 149-157.
- Littlepage, G., & Pineault, T. (1978). Verbal, facial, and paralinguistic cues to the detection of truth and lying. Personality and Social Psychology Bulletin, 4(3), 461-464.

- Littlepage, G.E., & Pineault, M.A. (1979). Detection of deceptive factual statements from the body and the face. Personality and Social Psychology Bulletin, 5(3), 325-328.
- Littlepage, G.E., & Pineault, M.A. (1981). Detection of truthful and deceptive interpersonal communications across information transmission modes. The Journal of Social Psychology, 114, 57-68.
- Littlepage, G.E., & Pineault, M.A. (1982). Detection of deception of planned and spontaneous communications. Unpublished manuscript, Middle Tennessee State University.
- Littlepage, G.E., McKinnie, R., & Pineault, M.A. (1983). Relationship between nonverbal sensitivities and detection of deception. Perceptual and Motor Skills, 57, 651-657.
- Lykken, D.T. (1978). The psychopath and the lie detector. Psychophysiology, 15(2), 137-142.
- Lykken, D.T. (1979). The detection of deception. Psychological Bulletin, 86(1), 47-53.
- Lykken, D.T. (1981). The lie detector and the law. Criminal Defense, 8, 19-27.
- Maier, R.A., & Lavrakas, P.J. (1976). Lying behavior and evaluation of lies. Perceptual and Motor Skills, 42, 575-581.
- Maier, N.R.F. (1966). Sensitivity to attempts at deception in an interview situation. Personnel Psychology, 19, 55-66.

- Maier, N.R.F., & Janzen, J.C. (1967). Reliability of reasons used in making judgments of honesty and dishonesty. Perceptual and Motor Skills, 25, 141-151.
- Maier, N.R.F., & Thurber, J.A. (1968). Accuracy of judgments of deception when an interview is watched, heard, and read. Personnel Psychology, 21, 23-30.
- Marston, W.M. (1917). Systolic blood pressure changes in deception. Journal of Experimental Physics, 2, 117-163.
- Matarazzo, J.D., Wiens, A.N., Jackson, R.H., & Manaugh, T.S. (1970). Interviewee speech behavior under conditions of endogenously-present and exogenously-induced motivational states. Journal of Clinical Psychology, 26, 141-148.
- Mehrabian, A. (1971). Nonverbal betrayal of feeling. Journal of Experimental Research in Personality, 5, 64-73.
- Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. Journal of Personality and Social Psychology, 6(1), 109-114.
- Miller, G.R., deTurck, M.A., & Kalbfleisch, P.J. (1983). Self-monitoring, rehearsal, and deceptive communication. Human Communication Research, 10(1), 97-117.
- Miller, G.R., & Kalbfleisch, P.J. (1982, May). Effect of self-monitoring and opportunity to rehearse on deceptive success. Paper presented at the convention of the International Communication Association, Boston, Mass.

- Mongeau, P. (1984, March). Personal Communication.
- Motley, M.T. (1974). Acoustic correlates of lies. Western Speech, Spr, 81-87.
- Olson, C.T. (1978). The effect of perceived conditions of interpersonal observation on encoding and decoding processes during deceitful self-presentations. Unpublished Doctoral Dissertation, Columbia University.
- Parker, R.J. (1978). Age, sex, and the ability to detect deception through nonvocal cues. Unpublished Doctoral Dissertation, Fresno Campus, California School of Professional Psychology.
- Podlesny, J.A., & Raskin, D.C. (1977). Psychophysiological measures and the detection of deception. Psychological Bulletin, 84, 782-799.
- Potamkin, G.G. (1982). Heron addicts and nonaddicts: The use and detection of nonverbal deception clues. Unpublished doctoral dissertation, California School of Professional Psychology.
- Raskin, D.C. (1978). Scientific assessment of the accuracy of detection of deception: A reply to Lykken. Psychophysiology, 15(2), 143-147.
- Raskin, D.C., & Hare, R.D. (1978). Psychopathy and detection of deception in a prison population. Psychophysiology, 15(2), 126-136.
- Raskin, D.C., & Podlesny, J.A. (1979). Truth and deception: A reply to Lykken. Psychological Bulletin, 86(1), 54-59.

- Raskin, D.C., & Podlesny, J.A. (1978). Effectiveness of techniques and physiological measures in the detection of deception. Psychophysiology, 15(4), 344-359.
- Reid, J.E., & Inbau, F.E. (1977). Truth and deception: The polygraph ("lie detection") technique (2nd ed.). Baltimore: Williams & Wilkins.
- Riggio, R.E., & Friedman, H.S. (1983). Individual differences and cues to deception. Journal of Personality and Social Psychology, 45(4), 899-915.
- Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.
- Rotkin, H.G. (1980). Information used in detecting deception. Unpublished doctoral dissertation, New York University.
- Rovira, M.L. (1982). Detection of deception: A signal detection theory analysis. Unpublished doctoral dissertation, The Catholic University of America.
- Sakai, D.J. (1981). Nonverbal communication in the detection of deception among women and men. Unpublished doctoral dissertation, University of California, Davis.
- Sereno, T.J.P. (1981). Children's honesty revisited: An exploration of deceptive communication in preschoolers. Unpublished doctoral dissertation, Bowling Green State University.

- Slowik, S.M., & Buckley, J.P. (1975). Relative accuracy of polygraph examiner diagnoses from respiration, blood pressure, and GSR recordings. Journal of Police Science and Administration, 3, 300-309.
- Stiff, J.B., & Miller, G.R. (1984). Interrogation, level of message exposure, and judgments of honest and deceit: Toward a more interactive model of communication. Paper presented at the Annual Western Speech Communication Association meeting, Seattle.
- Streeter, L.A., Krauss, R.M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. Journal of Personality and Social Psychology, 35(5), 345-350.
- Summers, W.G. (1939). Science can get the confession. Fordham Law Review, 8, 355-354.
- Snyder, M. (1974). Self-monitoring of expressive behavior. Journal of Personality and Social Psychology, 30, 526-537.
- Szucko, J.J., & Kleinmuntz, B. (1981). Statistical versus clinical lie detection. American Psychologist, 36, 488-496.
- Trovillo, P.V. (1939a). A history of lie detection. The Journal of Criminal Law and Criminology, 29(6), 848-875.
- Trovillo, P.V. (1939b). A history of lie detection. The Journal of Criminal Law and Criminology, 30(1), 104-119.

- Wicklander, D., & Hunter, F. (1975). The influence of auxiliary sources of information in polygraph diagnosis. Journal of Police Science and Administration, 3, 405-409.
- Wilson, S.J. (1975). Channel differences in the detection of deception. Unpublished doctoral dissertation, Florida State University. (University Microfilms No. 76-2726)
- Yohman, J.R. (1978). The guilty knowledge technique in lie detection. Biological Psychology Bulletin, 5(3), 96-103.
- Zuckerman, M., Amidon, M.D., Bishop, S.E., & Pomrantz, S.D. (1982). Face and tone of voice in the communication of deception. Journal of Personality and Social Psychology, 43(2), 347-357.
- Zuckerman, M., DePaulo, B.M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. Advances in Experimental Social Psychology, 14, 1-49.
- Zuckerman, M., Kernis, M.R., Driver, R., & Koestner, R. (1984). Segmentation of behavior: Effects of actual deception and expected deception. Journal of Personality and Social Psychology, 46, 1173-1182.
- Zuckerman, M., Koestner, R., & Alton, A. O. (1984). Learning to detect deception. Journal of Personality and Social Psychology, 35(5), 345-350.

Zuckerman, M., Koestner, R., & Colella, M. J. (in press) Learning
to detect deception from three communication channels.
Journal of Nonverbal Behavior.

Footnotes

¹The studies reviewed here required persons to observe communicators both lying and telling the truth at equal frequencies.

²Atmiyanandana, 1976; Bauchner, Brandt, & Miller, 1977; Brandt, Miller, & Hocking, 1980a; Brandt, Miller, & Hocking, 1980b; Brandt, Miller, & Hocking, 1982; Ekman & Friesen, 1974; Fay & Middleton, 1941; Harrison, Hwalek, Raney, & Fritz, 1978; Hemsley, 1977; Hemsley & Doob, 1979; Hocking, Bauchner, Kaminski, & Miller, 1979; Lavrakas, 1977; Littlepage & Pineault, 1979; Littlepage & Pineault, 1981; Littlepage & Pineault, 1982; Littlepage, McKinnie, & Pineault, 1983; Maier & Janzen, 1965; Maier & Lavrakas, 1976; Maier & Thurber, 1968; Matarazzo, Wiens, Jackson, & Manaugh, 1970; Miller, deTurck, & Kalbfleisch, 1983; Motley, 1974; Parker, 1978; Potamkin, 1982; Rovira, 1988; Sakai, 1981; Sereno, 1981; Stiff & Miller, 1984; Wilson, 1975; Zuckerman, Koestner, & Alton, 1984; Zuckerman, Kernis, Driver, & Koestner, 1984; Zuckerman, Koestner, & Colella, in Press.

³Atmiyanandana, 1976; Bauchner, Brandt, & Miller, 1977; Brandt, Miller, & Hocking, 1980a; Brandt, Miller, & Hocking, 1980b; Brandt, Miller, & Hocking, 1982; Comadena, 1981; DePaulo, Davis, & Lanier, 1980; DePaulo & Jordan, 1982; DePaulo, Jordan, Irvine, & Laser, 1982; DePaulo, Lanier, & Davis, Submitted; DePaulo, Lassiter, & Stone, 1982; DePaulo & Rosenthal, 1979;

DePaulo, Rosenthal, Green, & Rosenkrantz, in Press; DePaulo, Rosenthal, Rosenkrantz, & Green, Submitted; DePaulo, Stone, & Lassiter, 1984; Dollinger, Reader, Marnett, & Tylanda, 1983; Ekman & Friesen, 1969; Ekman & Friesen, 1974; Exline, Thibaut, Hickey, & Gumpert, 1970; Fay & Middleton, 1941; Feldman, 1976; Feldman, 1979; Feldman, Jenkins, & Popoola, 1979; Feldman & White, 1980; Fugita, Hogrebe, & Wexley, 1980; Geis & Moon, 1981; Geizer, Rarick, & Soldow, 1977; Harrison, Hwalek, Raney, & Fritz, 1978; Hemsley, 1977; Hemsley & Doob, 1978; Hemsley & Doob, 1979; Hildreth, 1953; Hocking, Bauchner, Kaminski, & Miller, 1979; Kraut, 1978; Kraut & Lewis, 1984; Kraut & Poe, 1980; Lavrakas, 1977; Littlepage & Pineault, 1978; Littlepage & Pineault, 1979; Littlepage & Pineault, 1981; Littlepage & Pineault, 1982; Littlepage, McKinnie, & Pineault, 1983; Maier, 1966; Maier & Janzen, 1965; Maier & Lavrakas, 1976; Maier & Thurber, 1968; Matarazzo, Wiens, Jackson, & Manaugh, 1970; Miller, deTurck, & Kalbfleisch, 1983; Motley, 1974; Olson, 1978; Parker, 1978; Pollman, 1982; Potamkin, 1982; Rotkin, 1980; Rovira, 1988; Sakai, 1981; Sereno, 1981; Stiff & Miller, 1984; Streeter, Krauss, Geller, Olson, & Apple, 1977; Wilson, 1975; Zuckerman, Amidon, Bishop, & Pomrantz, 1982; Zuckerman, Koestner, & Alton, 1984; Zuckerman, Kernis, Driver, & Koestner, 1984; Zuckerman, Koestner, & Colella, in Press.

⁴Mean accuracy ratings for study 11 are the combined ratings from both factual and emotional lies/truths. Studies 6, 14, and 2 do not have audio.

⁵Mean accuracy ratings for study 11 are the combined ratings from both factual and emotional lies/truths.

⁶Mean accuracy ratings for study 11 are the combined ratings from both factual and emotional lies/truths. The combined head and body with full audio observation condition for study 2 represents accuracy estimates made by subjects viewing and listening to stimulus subjects through a two way mirror.

MICHIGAN STATE UNIV. LIBRARIES



31293106824182