Polygraph

VOLUME 33 2004 NUMBER 2

Contents

Published Quarterly

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

Polygraph

Editor-in-Chief: Stuart M. Senter, Ph.D.

******

Associate Editors: Norman Ansley, Troy Brown, Ph.D., Andrew Dollins, Ph.D., Kim English, Frank Horvath, Ph.D., Murray Kleiner, Donald Krapohl, Vance MacLaren, Dean Pollina, Ph.D., Michael Sakuma, Ph.D., Shirley Sturm, Douglas Vakoch, Ph.D., Gordon L. Vaughan, Esq., Jennifer Vendemia, Ph.D., Virgil Williams, Ph.D., Lee Wurm, Ph.D.

Subscription Information: *Polygraph* is published quarterly by the American Polygraph Association. Advertising and Editorial Address is P.O. Box 10342, Ft. Jackson, SC 29207 (USA). Subscription Rates: One year $80.00 (domestic), $100.00 (foreign). Change of address: APA National Office, P.O. Box 8037, Chattanooga, TN 37414-0037. THE PUBLICATION OF AN ARTICLE IN POLYGRAPH DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT BY THE AMERICAN POLYGRAPH ASSOCIATION.

The Sheridan Press

Criminalistics Use of Polygraph in Russia: Modern Level and Development Prospects[1]

Yuri I. Kholodny[2] and P. V. Akentiev[3]

2003 is a jubilee year for polygraph specialists and those whose life is in some extent connected with polygraph. The year 1993 became the end of decades of incogitant denial of possible use of the "lie detection" psycho-physiological technique for law-enforcement purposes. Ten years ago, the Ministry of Justice enacted the first legal mandate in the country, which regulated polygraph use by the Federal Security Service thereby legalizing applied use of the technique in Russia. What are the achievements of national science and practice in this field for the last 10 years? 1994 saw examinations using polygraph pass into service in the Russian Ministry of Internal Affairs. In 1996, tax police began implementation of polygraph in its activities. In 1998, Ministry of Justice and Ministry of Defense did the same. Divisions of military prosecutor's office have recently started using polygraph examinations more often.

One may positively state that more than ten thousand checks with various purposes are completed annually using polygraph. Polygraph is increasingly widely used by commercial structures, and, according to rough estimations, the number of tests using this instrument in commerce exceeds that in the public sector. It is worth saying a good word about our engineers and programmers of public and commercial organizations, who set up production of computer polygraphs and made it possible to spare hundreds of thousands of dollars for our country due to stopping imports of this kind of instrument, though, some hotheads insisted on it in the middle of 1990s.

In general, there is progress evident in the development of examinations using polygraph in Russia within the last years. At the same time, the pathway effect of the long-term nihilism in connection with polygraph, which remains in our legal science and public consciousness, played a noticeable part in the slowdown of spreading this efficient technique in this country and appearance of several backlashes. Theoretical (to be more exact - natural scientific and legal, arising from it) justification of this technique is of vital importance in the activity of MIA, FSB and other federal authorities. Criminal science for several years failed to pay attention to the wide spread use of polygraph in internal affairs authorities and in the whole country, which began in 1993-1994, though the criminal science is inter alia to establish scientific justification and reliability of techniques used for the crime control. It was only in the late 1990s when criminal investigation science paid attention to this technique and, being unaware what to do with it, classified it as "an unconventional method of obtaining information necessary for crime investigation".

---

[1]The article has been reprinted with permission from the *Russian investigator* magazine.

[2]Doctor of Juridical Science, Candidate of Psychology, Institute of Criminalistics of Research and Technology Department of Russian FSB. Contact the author at: kholodny@ipc.ru

[3]Candidate of Medicine, ARISR of Russian MIA

Therefore, a paradoxical situation occurred by the beginning of the new century - a new method was increasingly spreading in law-enforcement practice, though this method had not been studied by criminal science while the so-called polygraph problem, which appeared as far back as in 1970s and covered a wide amorphous aggregate of methodical, technical, legal and other questions, actually remained unelaborated.

However, it became possible within the last two or three years to arrive at considerable success in understanding "the polygraph problem" and resolve a range of principal technical problems, which led to important practical consequences. Let us briefly characterize the level achieved in the field of theory and practice of examinations using polygraph early in the 21st century.

1. The modern world science and practice collected much practical scientific knowledge about the psycho-physiological technique of "lie detection", which was combined under the name criminalistic polygraphology. Analysis of available disembodied and scientific data shows that the psycho-physiological technique of "lie detection" is only one of the four methods which allow for the discovery of information concealed by a person, while polygraph examination is just one of the five particular forms of psycho-physiological technique realization.

2. When speaking about polygraph examinations, many specialists see it as a diagnostic method using emotional or psychical stress. Theoretical and experimental research of our scientists established psychological and neurophysiological mechanisms underlying the process of revelation of information concealed by a person. It was proved that from the natural scientific point of view, examination using polygraph is a complex human memory (not emotional/psychical stress, which is often present during polygraph examination) research technique (to be more exact - testing technique). This result is of vital importance for understanding the place and role of examination techniques using polygraph in the system of modern criminalistics.

3. Studying of polygraph examinations from the point of view of categorical criminalistics (taking into account comprehension of natural scientific fundamentals of its technology) made it possible to prove that the use of polygraph is the first to allow objective detecting and studying of events kept in the human memory, which events took place earlier in the person's life, including criminally relevant events, or revealing absence of such events. At the same time, it was proved that traces of events kept in the emotional memory of a person practically cannot be eliminated through the entire life of such a person. Polygraph examinations confirmed the previous statement: real practice successfully revealed traces of events in the human memory, which happened 20 or more years ago.

4. Evaluation of polygraph examination from the point of view of criminalistics methodology made it possible to prove that it is an individual technique of criminalistic diagnostics. Inclusion of polygraph examination and criminalistic polygraphology in modern criminalistics resulted in the appearance of quite a new trend in criminalistics that had not existed before, which studied apparatus (that is objective) diagnostics of presence (or absence) of event traces in the human memory, which events are important for crime detection and investigation.

5. Evaluation of the importance of polygraph technology for the purpose of detecting information concealed by a person in the system of modern national criminalistics afforded grounds for separating criminalistic polygraphology as a new independent criminalistic technique. This allowed to finally exclude polygraph examination from the range of "unconventional facilities for obtaining information important for crime investigation".

6. Polygraph use analysis from the point of view of criminalistic theory made it possible to justify and prove that polygraph examination is an individual technique of criminalistic prevention of crimes and offences connected with the course of duties. This conclusion is of vital importance in the light of the fact that screening examinations are currently introduced as a system preventive measure for

personnel selection in Russian MIA and FSB bodies.

7. Based on an understanding of events underlying the technology of polygraph detection of information concealed by a person and taking into account procedural regulations, we concluded that:

- firstly, polygraph use for examination currently seems to be rather questionable due to the absence of direct indications in the criminal procedure and, therefore, inexpedient;

- secondly, polygraph examination may be used in such investigative action as "identification parade" as means of deliberate non-identification;

- thirdly, the most promising form of procedural polygraph use is expertise aimed at evaluation of reliability of evidence given by the examined person in interrogation, investigation or before the court.

Theoretical grounds have been laid for a new type of expertise - legal and psychophysical expertise, and the first experience has been gathered in using such grounds for crime investigation. Another motive for polygraph implementation in law-enforcement activities was the inclusion of clause 3.1 in section 2 of art. 74 "Evidence" of the RF Code of Criminal Procedure on recognition of specialist's conclusion and testimony as acceptable evidence'.

8. Evaluation of legal regulation of polygraph use in the MIA and FSB bodies and other executive authorities shows that the departmental instructions in general reflect internationally accepted principles of using this technique. Alongside with that, these regulations have a few shortcomings and omissions, which must be removed. Taking into account the evident recent progress in the solution of the "polygraph problem", it becomes evident that legal regulation of polygraph use requires further improvement both on the departmental level and through adoption of the relevant federal act.

To sum up the information presented above, we may confidently state that the polygraph examination technique blends with the technique and device system of modern criminalistics, and it goes beyond the framework of operational search actions and thrusts its way to the forensic examination system.

Quality polygraph use for law-enforcement purposes depends on the four components: a) availability of a clear regulatory management system for applied polygraph use, b) availability of science-based methodical regulations of practical polygraph use, c) availability of reliable firmware (computer polygraphs), which pass the relevant certification, d) availability of qualified polygraph specialists.

The last component is the most important one due to the fact that 90% of polygraph examination efficiency depends on the polygraph specialist qualification.

It is possible to separate the most important current tasks of criminalistic polygraphology due to both the urgent necessity of wide implementation of polygraph in law-enforcement practice and the achieved level in theory and practice of examinations using this firmware. Task one is to improve the skills of polygraph specialists and eliminate departmental differences in the quality of their qualifications. An important condition of solving this task is the necessity to overcome separation and disunity of specialists from different authorities using polygraph. To solve this task, the Institute of Criminalistics of the Russian FSB, which is the centre of polygraph specialists training, entered into a scientific and technological cooperation agreement with the All-Russian Institute of Scientific Research of the Russian MIA and rendered consulting and methodical assistance in the organization of polygraph specialists courses jointly held by General Office for Special Technical Measures and ARISR of the Russian MIA. Specialists of IC of the Russian FSB take part in the united qualification and methodology committee of GOSTM and ARISR of the Russian MIA.

To successfully solve the first task, it is necessary to adjust to a common standard the technology of training of polygraph specialists for different federal authorities. This technology must comply with the world

Truth or Just Bias: The Treatment of the Psychophysiological Detection of Deception in Introductory Psychology Textbooks[1]

Mary K. Devitt, Charles R. Honts, and Lynelle Vondergeest

Abstract

This study examined the presentation of psychophysiological detection of deception (PDD; polygraph) testing in introductory psychology textbooks. We examined a sample of 37 introductory psychology textbooks published between 1987 and 1994 for content that discussed PDD testing. Excerpts concerning PDD were then checked for misdescriptions or inaccuracies and rated by two psychophysiologists and a social psychologist. The results showed that PDD received strongly negative treatment in the texts. Moreover, the treatments were often fraught with misdescriptions and inaccuracies. In addition, there was an over-reliance on reviews as opposed to empirical studies. We discuss the significance of the problems of bias, reliance on secondary sources, and inaccuracies, and elaborated on the importance of balanced and error free presentations in this medium that serves as a first introduction to the science of psychology for so many people.

Previous content analyses of Introductory Psychology textbooks have been conducted in areas such as the treatment of counseling versus clinical psychology (Leong & Poynter, 1991), transactional analysis (Douglass, 1990), humanistic psychology (Churchill, 1988), sensory deprivation research (Suedfeld & Coren, 1989), religion (Lehr & Spilka, 1989), parapsychology (Roig, Icochea, & Cuzzucoli, 1991), the number of neurons in the brain (Soper & Rosenthal, 1988), the Little Albert legend (Paul & Blumenthal, 1989), the Yerkes-Dodson Law (Winton, 1987), the utility of idealized figures (Shepard, 1983), and racial diversity (Gay, 1988). Those studies have illustrated that misdescriptions, inaccuracies, theoretical biases, ambiguity, lack of objectivity, or lack of assimilation may be present in Introductory Psychology material. As a result, it appears that college students are not being well served when controversial material is inadequately and incompletely presented.

In the present study we address another controversial area that is frequently covered in introductory textbooks, that is, the psychophysiological detection of deception (PDD). Psychophysiological detection of deception tests (also known as polygraph or lie detector tests) are psychological tests that are an important application of psychology in the real world (Honts, 1994a). In the United States and Canada, virtually all federal and local law enforcement agencies employ polygraph examiners who conduct investigative examinations with criminal suspects. The results of such tests often remove individuals from suspicion or result in confessions of wrongdoing following interrogations (Honts & Perry, 1992; Lykken, 1981; Raskin, 1986). Polygraph testing also finds application in the

[1] This article is reprinted with permission from *The Journal of Credibility Assessment and Witness Psychology* taken from Volume 1(1), pages 9 -32. Please refer to http://truth.boisestate.edu for copies for reproduction under the original copyright that follows:

This article was edited by J. Peter Rosenfeld of Northwestern University.

workplace (Honts, 1991). Although many screening uses of polygraph testing in the private sector were prohibited in 1988 (Employee Polygraph Protection Act), employers may still use polygraph to investigate specific losses, and several industries were exempted from the screening ban. Moreover, polygraph tests for pre-employment screening are widely used by federal, state, and local governments. Polygraph pre-employment screening of police officer applicants is particularly pervasive. Finally, polygraph testing plays a critical role in personnel selection and the security clearance process in the national security agencies (Department of Defense, 1991; Honts, 1991; 1994a). All employees of the National Security Agency and the Central Intelligence Agency must take and pass polygraph tests to obtain and retain their security clearances. There are proposals to expand greatly the numbers of individuals subject to such clearance testing (Department of Defense, 1991). Although the numbers of tests conducted in the national security system may be relatively small in absolute terms (i.e., in the tens of thousands), in terms of the special trust and power placed in the hands of those who conduct PDD examinations, the importance of such tests can hardly be overstated (Honts, 1994a). It thus seems important that Introductory Psychology textbooks present a fair and unbiased picture of this important area of applied psychology.

Method

Materials. The data base for this study consisted of an exhaustive sample of the 37 Introductory Psychology textbooks offered to the psychology faculty of a medium sized midwestern university during the academic year 1993/94. In the case that multiple editions of any textbook were made available, only the most current edition was used in this analysis.

Procedure. Each of the textbooks was searched for references to lie detection, polygraph, or detection of deception testing. If the textbooks contained references to PDD testing, the words and number of pages devoted to the topic were counted. The textbook sections were then rated on a 7-point scale regarding their orientation toward PDD testing (1 = negative, 4 = neutral, 7 = positive). Three different individuals rated the orientation for all of the textbook excerpts. The first rater was a psychophysiologist (the second author of the present manuscript) who was highly familiar with the polygraph testing literature and who has testified as an expert on polygraph examinations in a number of courts of law in the United States and Canada. The second was an assistant professor of psychology (a colleague of the first author) who was trained as a psychophysiologist and who was not involved in polygraph research or in the polygraph controversy in any way. The third evaluator was an associate professor social psychologist (a colleague of the second author) who has not been involved in the polygraph controversy in any way, but who does frequently teach large Introductory Psychology classes. The three evaluations were conducted independently of one another.

In addition, reference citations were recorded. The reference citations present in each textbook were counted, examined, and classified as to their orientation (either positive or negative) toward polygraph testing. The reference citations were also classified as either laboratory or field studies, or reviews. When research or reviews were cited, the descriptions of empirical research and reviews were examined for factual errors or misdescriptions. Also recorded were the types of polygraph usage (forensic testing, investigative testing, on-the-job screening, pre-employment screening, or national security screening) discussed in each textbook. The types of polygraph tests (Control Question Test, Concealed Knowledge Test, or Relevant-Irrelevant Test) mentioned were also recorded.

Results

General Statistics. The data collected in this study are presented and summarized in Table 1. The mean textbook length was 656.31 pages. For only those books that discussed polygraph testing, the mean textbook length was 655.9 pages. The mean number of pages devoted to a discussion of polygraph testing was 1.5 pages. Twenty-nine of the textbooks (78.4%) included some discussion of polygraph testing. Of the texts that discussed PDD, only 11 (29.7%) described empirical research.

Ratings. The mean ratings of the textbook excerpts were as follows: Polygraph-Expert/Psychophysiologist, M = 2.24, sd = 0.87, Independent Psychophysiologist, M = 2.55, sd =1.01, Social Psychologist, M = 3.79, sd =0.86. A repeated measures Analysis of Variance (ANOVA) was used to test for differences among the raters. This analysis revealed a significant difference between the means, $F(2, 27) = 40.51$, $p < .001$. This analysis was followed-up with single degree of freedom tests. The Bernoulli corrected p value was calculated by dividing alpha (.05) by the number post-hoc comparisons (3), for an alpha value of $p = .017$. The univariate tests indicated that the ratings by the two psychophysiologists were not significantly different, $F(1, 28) = 1.95$, $p > .1$, but that the ratings of the Polygraph-Expert Psychophysiologist and the Social Psychologist were significantly different, $F(1, 28) = 71.95$, $p < .001$, as were the ratings of the Independent Psychophysiologist and the Social Psychologist, $F(1, 28) = 37.57$, $p < .001$. Interestingly, neither of the psychophysiologists rated a single excerpt as positive. The average rating for the three evaluators is shown in Table 1.

Citations. Only four (16%) of the textbooks provided any positive citations, and those textbooks cited only review articles. The ratio of negative citations to positive citations was over 15 to 1 (4.28/.28). To determine if differences in orientation, number, and type of citations existed between the textbooks that discussed both empirical studies and reviews (Mixed group) and those textbooks that discussed reviews only (Review group), t-tests for independent samples were conducted. There was a significant difference in orientation for the discussion type, $t(23) = 3.23$, $p = .003$, with the Mixed group providing a more negative discussion (M = 1.64) than the Review group (M = 2.64). Also noted was a significant difference in total number of citations provided by each group, $t(23) = 3.28$, $p = .003$. The Mixed group provided more citations (M = 6.46) than the Review group (M = 3.0). Finally, there was a significant difference in the number of negative citations $t(23) = 3.96$, $p = .001$, with the Mixed group presenting more negative citations (M = 6.46) than the Review group (M = 2.57).

Table 1. *Analysis of the Presentation of Polygraph Testing in 37 Introductory Psychology Texts*

Text (First Author Shown)	**Number of Words on PDD**	**Number of Negative Cites**	**Number of Positive Cites**	**Average Rating (1 = Negative)**
Segments with Both Empirical and Reviews Cited				
Atkinson	1,223	3	0	3.67
Doyle	520	5	0	3.00
Dworetzky	1,899	12	0	1.33
Feldman	625	10	0	2.33
Huffman	587	8	0	2.33
Kalat	875	7	0	3.00
Lefton	502	4	0	2.67
Santrock	722	7	0	3.00
Wade	764	3	0	2.33
Wood	959	6	0	1.33
Worchel	1,178	6	0	3.00
Means	896	6.5	0	2.54
Only reviews Cited				
Baron	606	2	0	2.67

Table 1. Continued

No Citations				
Bernstein	575	2	2	3.33
Bootzin	274	2	0	2.33
Carlson	1,135	2	0	3.67
Gleitman	301	2	2	3.33
Laird	584	3	0	1.67
Meyers	1,281	9	1	3.00
Peterson	312	1	0	3.00
Pettijohn	232	1	0	3.67
Rubin	182	2	0	2.67
Smith	834	5	2	2.67
Weiten	520	23	0	3.00
Weiten (briefer version)	531	2	0	3.00
Wortman	541	1	0	4.33
Means	565	2.6	0.5	3.02
Crooks	673			2.33
Darley	252			3.00
Roediger	182			4.00
Shaver	167			3.33
Means	318			3.17
PDD Not Discussed				
Benjamin				
Bourne				
Gerow				
Goldstein				
Gray				
McConnell				
Ornstein				
Zimbardo				
Grand Means	656	4.3	0.3	2.86

The frequency of various citations was also examined. The most frequently cited (14 times) review was the popular book by Lykken (1981). The most commonly mentioned (6 times) empirical field study was one by Kleinmuntz and Szucko (1984). Finally, the most commonly cited (2 times each) laboratory validity studies were the studies by Honts, Hodes, and Raskin (1985; concerning countermeasures) and by Szucko and Kleinmuntz (1981; concerning validity). Overall, 64 different citations were noted. Fifty (78.1%) of those citations were for reviews, nine (14.1%) were empirical laboratory studies, and five (7.8%) were empirical field studies. Fifteen reviews, two laboratory studies, and three field studies were cited more than one time each. Furthermore, over all of the textbook excerpts there were 113 citations (i.e., some of the 64 separate citations were cited in more than one textbook). The most

frequently cited author was David Lykken with a total of 29 citations for eight different publications. At least one of Lykken's works was discussed in 19 of the textbooks.

The types and uses of polygraph testing discussed in the excerpts were also assessed. Those results are presented in Table 2. Overall, 23 of the textbooks discussed some specific use or type of polygraph tests. In those texts that discussed types of polygraph testing (Control Question Test [CQT], Concealed Knowledge Test [CKT], and Relevant/Irrelevant [RI]), 17 (74%) mentioned only one test type. The other six textbooks mentioned two types of polygraph tests. No textbooks discussed more than two types of polygraph tests. Ten textbooks provided a discussion of the RI test. The CQT and the CKT were each discussed in nine textbooks. The uses of polygraph tests that were assessed included forensic testing, investigative testing, on-the-job screening, pre-employment screening, and national security screening. Overall, 23 (62.1%) of the textbooks included some mention of at least one of the uses of polygraph tests, although only the textbooks with citations (reviews and/or empirical research) discussed those uses. Pre-employment screening was discussed most often (17 times), followed by forensic and investigative testing (14 times each). On-the-job screening was mentioned 13 times, and national security screening was discussed in eight textbooks. Only one textbook discussed all of the polygraph uses. Thirteen textbooks discussed three or four of the uses, while nine textbooks discussed either one or two of the possible uses.

Finally, the discussions of polygraph testing were examined for factual errors in the reported research. Overall, 25 textbooks provided discussions with research or review citations. Factual errors or misdescriptions were noted in 18 (72%) of those textbooks (e.g., Feldman [1993], in describing a countermeasure study by Honts, Hodes, and Raskin [1985], stated that subjects in that study had used a tack in the shoe as a countermeasure. No such manipulation was included in that study). Details of the errors and misdescriptions in the excerpts are provided in Appendix A at the end of this article.

Discussion

Our analysis of the treatment of PDD in Introductory Psychology textbooks indicates that most textbooks present a negative view of the area. If the majority of research concerning PDD indicated poor validity, this view would clearly be justified. The question thus becomes what does the empirical literature have to say about the validity of PDD tests?

Table 2. *Percent of Textbooks That Provided a Discussion of the Uses and Types of Polygraph Tests*

Topic (Use/Type)	**Research & Reviews**	**Reviews Only**	**No Citations**
	(n = 11)	(n = 14)	(n = 4)
Forensic	34.6	64.3	00.0
Investigative	45.5	57.1	00.0
On-The-Job Screening	63.6	42.9	00.0
Pre-employment Screen	90.0	42.9	00.0
National Security Screen	45.5	21.4	00.0
Control Question Tests	36.4	35.7	00.0
Concealed Information	45.5	28.6	00.0
Relevant Irrelevant	27.3	35.7	50.0

A Brief Review of the Empirical Literature on PDD. Despite their widespread application, polygraph tests have been, and continue to be, the source of great controversy in the scientific literature. Of the three techniques discussed in this paper, there seems to be general agreement in the scientific literature that the Relevant-Irrelevant Test lacks validity (Ben-Shakhar & Furedy, 1990; Honts, 1991; Iacono & Patrick, 1988; Kleinmuntz & Szucko, 1982; Lykken 1981; Raskin, 1986; Saxe, Dougherty, & Cross, 1985). However, this may be a limited finding as the RI is used very infrequently in forensic settings and its applied uses seem to be limited to employment settings (Honts, 1991). If authors intend that their comments be directed to the use of the RI in employment settings they should state this clearly, as such incontrovertible agreement is noticeably lacking for the other two techniques.

The most commonly used test in the field is the Control Question Test. We will focus most of our analysis on validity studies of the CQT. The third technique, the Concealed Knowledge Test has been studied extensively in the laboratory, but has not achieved much application in the field. In the following section, we also review the empirical literature on the CKT.

The subsequent review also focuses on forensic applications of the polygraph. There is virtually no empirical scientific literature on the validity of PDD tests in employment settings, and thus there is nothing to review (Honts, 1991). Similarly there is little empirical literature on the national security uses of the polygraph. However, what literature there is on the national security uses consistently produces near chance estimates of validity (Barland, Honts, & Barger, 1989; Honts, 1991; 1992; 1994a). We found no references to any of theses sources in the Introductory Psychology textbooks.

Laboratory Studies Concerning Forensic Settings. A recent meta-analysis of 15 laboratory studies (Kircher, Horowitz, & Raskin, 1988) of the Control Question Test indicated a wide range of validity estimates. One study found near chance results, while six of the studies produced moderate validity estimates, and eight of the studies report validity coefficients of 0.7 or better. In four of the studies, the validity coefficients exceeded 0.8. The Kircher et al. meta-analysis noted that these laboratory studies differed widely in their ecological validity. Some studies used mock crimes and procedures that closely modeled field conditions while other studies were very artificial and used unrealistic procedures. Moreover, the Kircher et al., meta-analysis indicated that those laboratory studies that most closely modeled field conditions produced the highest accuracy rates. A similar state of affairs appears to exist in the Concealed Knowledge Test literature. A more recent review (Honts & Quick, 1995) of the most ecologically valid laboratory studies of both the CQT and the CKT produced overall estimates of accuracy of about 90% and approximately equal false positive and false negative error rates.

Regardless of their methodology, some (e.g., Ben-Shakhar & Furedy, 1990; Lykken, 1981) have criticized all laboratory studies on the grounds that they lack ecological validity. These critics contend that it is not possible in the laboratory to mimic adequately the motivational and emotional context of being given a polygraph test when you are accused of a crime. Others have argued that if sufficient care is taken in creating a deceptive context in the laboratory, then laboratory studies can be useful in estimating the accuracy of the technique in the field (e.g., Podlesny & Raskin, 1978; Kircher et al., 1988).

The Kircher et al. (1988) review and meta-analysis should have been easily available to all of the authors of the Introductory Psychology textbooks considered in this analysis. It was published in a first tier psychology journal (*Law and Human Behavior*) that is published by APA Division 41, and is abstracted in all of the popular reference sources. We believe that it is telling that the laboratory study cited most frequently for estimates of validity is the Szucko and Kleinmuntz (1981; *American Psychologist*) study which produced the lowest estimate of accuracy (detection efficiency $r = 0.21$; the next lowest study, which produced an r of .51, accounting for six times the criterion variance, is the Kircher et al., meta-analysis). Conspicuously absent from the textbook excerpts were references to equally available publications in first tier journals that produce high estimates for the validity of the Control Question Test (e.g., Podlesny & Raskin, 1978,

Psychophysiology; Ginton, Netzer, Elaad, & Ben-Shakhar, 1982, *Journal of Applied Psychology*; Kircher & Raskin, 1988; *Journal of Applied Psychology*; Dawson, 1981; *Psychophysiology*; Raskin & Hare, 1978; *Psychophysiology*). As a minimum, each of the studies cited above accounted for 10 times the criterion variance of Szucko and Kleinmuntz (the validity coefficient for Szucko and Klienmuntz was .21 while the validity coefficients for the cited studies ranged from .65 for Ginton et al., to .87 for Raskin and Hare). One is left with the inescapable conclusions that either the introductory psychology textbook authors gave only a cursory review to the laboratory data on the polygraph or they were biased in their choice of studies to cite.

Ben-Shakhar and Furedy (1990) provide a review of the laboratory studies of the Concealed Knowledge Test. At that time they found ten laboratory studies of the CKT that they felt were scientifically sound enough to include in their review (Balloun & Homes, 1979; Bradley & Ainsworth; 1984; Bradley & Warfield, 1984; Davidson, 1968; Giesen & Rollison, 1980; Lykken, 1959; Podlesny & Raskin, 1978; Steller, Haenert, & Eiselt, 1987; Stern, Breen, Watanabe, & Perry, 1981; Waid, Orne, Cook, & Orne, 1978). However, no meta-analysis or quantitative analysis of the quality of these studies was reported. Over all ten studies, the accuracy with guilty subjects ranged from 61.1% (Balloun & Holmes, 1979) to 100% (Bradley & Ainsworth, 1984; and Bradley & Warfield, 1984). Accuracy with innocent subjects ranged from 80.6% (Waid et al., 1978) to 100% in seven of the studies (Bradley & Ainsworth, 1984; Bradley & Warfield, 1984; Davidson, 1968; Giesen & Rollinson, 1980; Lykken, 1959; Podlesny & Raskin, 1978; Steller et al., 1987). Only a single one of these studies received a single citation in one textbook. That study was Bradley and Ainsworth (1984), one of two studies indicating 100% accuracy with both innocent and guilty subjects.

Field Studies Concerning Forensic Settings. In any event, laboratory studies cannot tell the complete story. Data from real world settings are necessary to compliment and extend the results from the laboratory. Unfortunately, validity estimates based on field studies are also mixed and highly debated. Much of the debate regarding field studies concerns the issue of what constitutes adequate methodology. There seems to be an emerging consensus among both proponents (e.g., Honts & Perry, 1992) and critics (e.g., Patrick & Iacono, 1991) that the following are the necessary minimum requirements for field studies of PDD: First, the subjects must represent the population for generalization. If one is interested in studying criminal suspects, then the subjects should be criminal suspects. Second, the cases used in the study should be selected by some random process without reference to the accuracy of the original examiners decision or to the quality of the physiological data. Third, the decisions used for the data analysis should be based on independent reviews of only the physiological data. Information about the case facts and the overt behavior of the subjects should be withheld from the evaluators. (This criterion holds only if the goal of the study is to determine the ability of the physiological data to discriminate the innocent and guilty. If the goal of the study is to determine the utility of the procedure for some applied goal, admissibility in court for example, the data from the original examiners may be more valuable, see Honts & Quick, 1995.) Fourth, the independent evaluators should be experienced in the independent evaluation of PDD data and they should use techniques that are representative of those actually used in the field. Finally, the truthfulness of the subjects must be confirmed by some criterion that is independent of the outcome of the polygraph examination. Confessions, although problematic, are generally considered to be the best criterion, especially if they are supported by corroborating evidence.

A recent review (Honts & Quick, 1995), found four field studies of the CQT (Honts, 1994b, now in press; Honts, & Raskin, 1988; Iacono & Patrick, 1991; and Raskin, Kircher, Honts, & Horowitz, 1988) and two of the CKT (Elaad, 1990; Elaad, Ginton, & Jungman, 1992) that were able to meet the stringent requirements for a useful field study described above. Three of the field studies (Honts, 1994; Honts & Raskin, 1988; Raskin et. al., 1988) produced accuracy rates above 90%. The independent evaluators in the third study (Iacono & Patrick, 1991) produced a high false positive rate,

although the accuracy rate of the original examiners exceeded 90%.

Recently Patrick and Iacono (1991) have suggested that retrospective field studies may not be useful for estimating the accuracy of polygraph tests because of sampling biases built into the design of such studies. Their position is based on a theoretical analysis and an earlier thought experiment (Iacono, 1991). Fortunately there is no compelling data to support their analysis and many of the assumptions of that analysis are insupportable (e.g., If a guilty person passes a polygraph test, there will be no further investigation of that suspect, and confessions are only obtained following failed polygraph tests). If these assumptions are altered or are invalid then very different conclusions can be suggested (Raskin, Honts, & Kircher, in press). Moreover, recent work contradicts their position (Honts, in press) and indicates that confession results are very comparable with results based on other criteria.

Unfortunately, only Iacono and Patrick (1991) would have been readily available to the authors of the Introductory Psychology textbooks considered here and it would have appeared in print as most of these texts would have been nearing completion. It is not fair to expect that the authors of Introductory Psychology textbooks should know about unpublished reports in an applied area. However, there were a number of other field studies that were available to these authors at the time these books were written. All of those studies were reviewed in a study commissioned by the United States Congress and conducted by the Office of Technology Assessment (OTA, 1983). The OTA report was subsequently summarized in the *American Psychologist* (Saxe, Dougherty, & Cross, 1985). OTA concluded that there were ten field studies of the Control Question Test that met minimal scientific standards (although none would unambiguously meet all of the criteria described above [Barland & Raskin, 1976; Bersh, 1969; Davidson, 1979; Horvath, 1977; Horvath & Reid, 1971; Hunter & Ash, 1973; Kleinmuntz & Szucko, 1982; Raskin, 1976; Slowik & Buckley, 1975; Wicklander & Hunter, 1975]). Over these ten studies, the average accuracy with guilty subjects was 90% and the average accuracy with innocent subjects was 80%. In those eight studies that used a confession criterion, the accuracy of decisions with guilty subjects ranged from 98.6% (Wicklander & Hunter, 1975) to 75% (Klienmuntz & Szucko, 1982). With innocent subjects the accuracy rates ranged from 100% (Davidson, 1979) to 51.1% (Horvath, 1977).

At present there are only two published field studies of the CKT. Both of those studies would meet the criteria described above for a useful field study of the detection of deception. The two studies were reported by Elaad and his colleagues (Elaad, 1990; Elaad, Ginton, & Jungman, 1992). The average accuracy rate for guilty subjects in those studies was 47% while the average accuracy with innocent subjects was 98%. These results suggest that in the field the CKT produces extremely high numbers of false negative errors. This finding has been discussed in the light of what we know about eyewitness memory, and may not be surprising (see the discussion in Raskin et al., in press).

Thus, like the laboratory studies, the high quality field studies also seem to paint a relatively positive picture of the accuracy of the CQT, although one could argue that the literature is mixed in both venues. The picture for the CKT is clearer, both the laboratory and the field studies indicate that the CKT is prone to false positive errors and that in field settings the false negative rate may be extreme.

Attitudes of the Scientific Community Toward PDD. Another index of the scientific community's view of PDD testing could be found in surveys. The members of the Society for Psychophysiological Research (SPR) were polled on this topic by The Gallup Organization (1984). At that time, 63% of the respondents said that they believed polygraph tests were useful diagnostic tools when used with other available information, while only 1% of the respondents stated a belief that polygraph tests were without value. More recently, the members of SPR were again surveyed about their attitude toward polygraph testing (Amato & Honts, 1994). The results of the Amato and Honts study showed that 60.2% of the respondents believed PDD tests were useful diagnostic tools when used with other available information. Moreover, 80.5 % of the respondents who claimed to be familiar with the PDD literature believed that

polygraph tests were useful diagnostic tools. Only 1.7% of the respondents stated that polygraph tests were without merit.

Discussion of the Present Results

Although there is controversy, the empirical and review literature concerning PDD suggests the following conclusions: There is little support for the Relevant-Irrelevant Test, but this test is in frequent use only in employment settings. The laboratory and field data concerning the Control Question Test are mixed. However, when the ecologically valid laboratory studies and the high quality field studies are considered, both indicate high validity for the CQT. The ecologically valid laboratory studies and the high quality field studies of the Concealed Knowledge Test converge on a conclusion that the CKT is prone to false negative errors. Moreover, in the field the CKT seems to produce extreme numbers of false negative errors.

Given the generally favorable findings of both the empirical laboratory and field literature on the CKT, our review of Introductory Psychology textbooks appears to have revealed a distressing lack of balance. None of the textbooks accurately noted the important distinctions in the literature concerning the validity of the three techniques. Moreover, the general negative tone of the textbooks appears to be unjustified by the literature. This lack of balance is typified by the fact that the most commonly cited field study of PDD was the study by Kleinmuntz and Szucko (1984). Of all the field studies available in the literature, regardless of quality, this study is the one of two confession studies (the other is Horvath, 1977) that produced notably lower accuracy estimates. Of the eight confession confirmation studies in the OTA report, these are the two with the worst accuracies.

Given that it is so frequently cited, it may be illustrative to describe the methodology of the Kleinmuntz and Szucko (1984) study at this time. Unfortunately, the most cited form of this study is a 1984 publication in the journal *Nature* which is only about one page in length. Very few details are provided in that publication. However, the study has been described in detail elsewhere (OTA, 1983, and in Kleinmuntz & Szucko, 1982). From those descriptions we can determine the following facts about the Kleinmuntz and Szucko (1984) study. The subjects of this study were individuals who were tested by a private company regarding employee theft as a condition of their employment. None of the subjects was under criminal investigation at the time of testing. The physiological data were evaluated by students of a polygraph school who had not completed their training. The polygraph school these students were attending is one that stresses the evaluation of the case facts and the subject's overt behavior. The independent quantitative analysis of the physiological data is not stressed. Finally, the student evaluators were given only 1/9th of the data they would usually have in making an evaluation and they were forced to use an unfamiliar rating scale with which they had no prior experience or training. That rating scale is never used in the field, and the students were not allowed to arrive at an inconclusive outcome, as they would be allowed to do in the field. The cases used were confirmed by confession, but the method of case selection was not specified in the report. There is no indication that any additional confirmatory information was sought or obtained. If the criteria for a useful field study described above are consulted, it can readily be seen that the Kleinmuntz and Szucko (1984) study fails on almost every count. However, none of these methodological shortcomings were mentioned by any of the Introductory Psychology textbook authors who referenced this study.

Another problematic field study that is frequently cited is one by Horvath (1977). One problem with that study is that the cases were selected for inclusion in the study on the basis of the quality of the recordings, not on some random sampling basis. Moreover, although it is not indicated in the *Journal of Applied Psychology* publication, the dissertation (Horvath, 1974) upon which it is based states that some of the innocent subjects were crime victims who were being tested to verify their statements to the police. Subsequent analyses indicated that all but one of the false positive errors occurred with innocent victims, not suspects (see Raskin, 1986).

We realize that the authors of Introductory Psychology textbooks do not have the time to read each dissertation upon which an empirical report is based, or to read all the available overlapping sources. However, the

critical information about the Kleinmuntz and Szucko (1984) and Horvath (1977) studies discussed above was available to the Introductory Psychology textbook authors discussed here through several published reviews (notably, Raskin 1986; 1987; 1989). The 1987 review by Raskin would have been readily revealed by even a cursory search on *PsycLit.*

Unfortunately, similar biases are evident in the descriptions of laboratory studies. One of the two most frequently cited laboratory studies (Szucko & Kleinmuntz, 1981) was the only study in the Kircher et al. (1988) meta-analysis that produced chance discrimination. As such, it was an extreme outlier in the negative direction. The other frequently cited laboratory study was by Honts et al. (1985). Although this study produced moderate discrimination rates in its control conditions, it was cited in the Introductory Psychology textbooks because it demonstrated that under certain circumstances PDD tests could be distorted and/or defeated by countermeasures. Thus, this article was also used to paint PDD testing in a negative light. Numerous laboratory studies published in readily available first tier journals were available to the Introductory Psychology textbook authors, but were ignored or overlooked in favor of an outlier in the negative direction.

Through their choice of citations, the authors of Introductory Psychology textbooks have painted a very negative picture of the science of PDD testing. Our review of the scientific literature shows that this extreme negative view is not justified. Although there is controversy, we strongly believe that the empirical literature supports the validity of polygraph testing with the Control Question Test. Moreover, scientific surveys indicate that the majority of psychophysiologists agree. We believe that most of the current treatments of PDD in Introductory Psychology textbooks are doing an injustice to newcomers to psychology by painting a distorted and biased view of this important applied psychology. At the worst, it could be argued that Introductory Psychology textbook authors should note that there is controversy and describe data from both sides. If studies such as Klienmuntz and Szucko (1984) are cited, the criticism of such studies should always be mentioned. Such a neutral position would seem to be defensible.

It would appear that Introductory Psychology textbook authors would do well to actually examine the research literature in controversial areas they write about, rather than relying on secondary sources that may have been written by extreme proponents for one side or the other in an ongoing controversy. Truth, rather than bias, should be the criterion for inclusion in this important format that introduces most people to scientific psychology.

Authors Notes

We wish to thank Eric Landrum and Marc Pratarelli for their assistance in reviewing and rating the textbook excerpts for this study. Thank you also to Peter Rosenfeld and the anonymous reviewers for their helpful comments and suggestions on earlier drafts of this article. Address correspondence to Mary Devitt, Department of Psychology, 215 N. Murray, Stillwater, OK 74078.

References

Amato, S. L., & Honts, C. R. (1994). What do psychophysiologists think about polygraph tests? A survey of the membership of SPR. *Psychophysiology, 31*, S22. (Abstract).

Atkinson, R. L., Atkinson, R. C., Smith, E. E., & Bem, D. J. (1993). *Introduction to psychology* (11th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Balloun, K. D., & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology, 64*, 316-322.

Barland, G. H., Honts, C. R., & Barger, S. D. (1989). *Studies of the Accuracy of Security Screening Polygraph Examinations.* Department of Defense Polygraph Institute, Fort McClellan, Alabama.

Barland, G. H., & Raskin, D. C. (1976). *Validity and reliability of polygraph examination of criminal suspects* (Contract No. 75-N1-99-0001). Washington D. C.: National Institute of Justice, Department of Justice.

Baron, R. A. (1992). *Psychology* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Benjamin, L. T., Hopkins, J. R., & Nation, J. R. (1990). *Psychology.* New York: Macmillan.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception.* New York: Springer-Verlag.

Bersh, P. J. (1969). A validation study of polygraph examiner judgments. *Journal of Applied Psychology, 53,* 399-403.

Bernstein, D. A., Clarke-Stewart, Roy, E. J., Srull, T. K., & Wickens, C. D. (1994). *Psychology* (2nd ed.). Boston: Houghton Mifflin.

Bootzin, R. R., Bower, G. H., Crocker, J., & Hall, E. (1991). *Psychology today: An introduction* (7th ed.). New York: McGraw-Hill.

Bourne, L. E., Jr., Ekstrand, B. R., & Dunn, W. L. S. (1988). *Psychology: A concise introduction.* New York: Holt, Rinehart, and Winston.

Bradley, M. T., & Ainsworth, D. (1984). Alcohol and psychophysiological detection of deception. *Psychophysiology, 21,* 63-71.

Bradley, M. T., & Warfield, J. F. (1984). Innocence, information, and the guilty knowledge test in the detection of deception. *Psychophysiology, 21,* 683-689.

Carlson, N. R. (1993). *Psychology: The science of behavior* (4th ed.). Needham Heights, MA: Allyn and Bacon.

Churchill, S. D. (1988). Humanistic psychology in introductory textbooks. *Humanistic Psychologist, 16,* 341-357.

Crooks, R., & Stein, J. (1991). *Psychology: Science, behavior, and life* (2nd ed.). Orlando, FL: Holt, Rinehart, and Winston.

Darley, J. M., Glucksberg, S., & Kinchla, R. A. (1991). *Psychology,* 5th ed. Englewood Cliffs, NJ: Prentice-Hall.

Davidson, P. O. (1968). Validity of the guilty knowledge technique: The effects of motivation. *Journal of Applied Psychology, 52,* 62-65.

Davidson, W. A. (1979). Validity and reliability of the cardio activity monitor. *Polygraph, 8,* 104-111.

Dawson, M. E. (1981). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers, *Psychophysiology, 17,* 8-17.

Department of Defense (1991). *Fiscal Year 1990 Report to Congress on the DoD Polygraph Program.* Washington, D. C.: Department of Defense.

Douglass, H. J. (1990). Transactional analysis in American college psychology textbooks. *Transactional Analysis Journal, 20*, 92-110.

Doyle, C. L. (1987). *Explorations in psychology.* Monterey, CA: Brooks/Cole.

Dworetzky, J. P. (1991). *Psychology* (4th ed.). St. Paul, MN: West.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology, 75*, 521-529.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology, 77*, 757-767.

Employee Polygraph Protection Act of 1988. Public Law 100-347, 29, U. S. C. 2001 (1988).

Feldman, R. S. (1993). *Understanding psychology* (3rd ed.). New York: McGraw-Hill.

The Gallup Organization (1984). Survey of members of the Society For Psychophysiological Research concerning their opinions of polygraph test interpretation. *Polygraph, 13*, 153-165.

Gay, J. (1988). The incidence of photographs of racial minorities in introductory psychology texts. *The Journal of Black Psychology, 15*, 77-79.

Gerow, J. R. (1992). *Psychology: An introduction* (3rd ed.). New York: Harper Collins.

Giesen, M., & Rollison, M. A. (1980). Guilty knowledge versus innocent associations: Effects of trait anxiety and stimulus context on skin conductance. *Journal of Research in Personality, 14*, 1-11.

Ginton, A., Netzer, D., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology, 67*, 131-137.

Gleitman, H. (1991). *Psychology* (3rd ed.). New York: W. W. Norton.

Goldstein, E. B. (1994). *Psychology.* Pacific Grove, CA: Brooks/Cole.

Gray, P. (1991). *Psychology.* New York: Worth.

Honts, C. R. (1991). The emperor's new clothes: Application of polygraph tests in the American workplace. *Forensic Reports, 4*, 91-116.

Honts, C. R., (1992). Counterintelligence scope polygraph (CSP) test found to be a poor discriminator. *Forensic Reports, 5*, 215-218.

Honts, C. R. (1994a). The psychophysiological detection of deception. *Current Directions in Psychological Science, 3*, 77-82.

Honts, C. R. (1994b). Field validity study of the Canadian Police College polygraph technique. *Psychophysiology, 31*, S57. (Abstract)

Honts, C. R. (1996). Criterion development and validity of the Control Question Test in field application. *The Journal of General Psychology, 123*, 309-324.

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70*, 177-187.

Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Challenges and changes. *Law and Human Behavior, 16,* 357-379.

Honts, C. R., & Quick , B. D. (1995). The polygraph in 1996: Progress in science and the law. *North Dakota Law Review, 71,* 997-1020.

Honts, C. R., & Raskin, D. C. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration, 16,* 56-61.

Horvath, F. A. (1974). *The accuracy and reliability of police polygraphic ("lie detectors") in examiners's judgments of truth and deception: The effect of selected variables.* Unpublished Doctoral Dissertation, Michigan State University.

Horvath, F. S. (1977). The effect of selected variables on 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. *The Journal of Criminal Law Criminology and Police Science, 62,* 276-281.

Huffman, K., Vernoy, M., Williams, B., & Vernoy, J. (1991). *Psychology in action* (2nd ed.). New York: Wiley.

Hunter, F. L., & Ash, P. (1973). The accuracy and consistency of polygraph examiners's diagnoses. *Journal of Police Science and Administration, 1,* 370-375.

Iacono, W. G., (1991). Can we determine the accuracy of polygraph tests? In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.) *Advances in psychophysiology* (Vol. 4). Greenwich, CT: JAI Press.

Iacono, W. G., & Patrick, C. J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.), *Clinical assessment of malingering and deception.* New York: Guilford. (205-233).

Kalat, J. W. (1993). *Introduction to psychology* (3rd ed.). Belmont, CA: Brooks/Cole.

Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12,* 79-90.

Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73,* 291-302.

Kleinmuntz, B., & Szucko, J. J. (1982). On the fallibility of lie detection. *Law and Society Review, 17,* 84-104.

Kleinmuntz, B., & Szucko, J. J. (1984). A field study of the fallibility of polygraphic lie detection. *Nature, 308,* 449-450.

Laird, & Thompson, (1992). *Psychology.* Boston: Houghton Mifflin.

Lefton, L. A. (1994). *Psychology* (5th ed.). Needham Heights, MA: Allyn and Bacon.

Lehr, E., & Spilka, B. (1989). Religion in the introductory psychology textbook: A comparison of three decades. *Journal for the Scientific Study of Religion, 28,* 366-371.

Leong, F. T., & Poynter, M. A. (1991). The representation of counseling versus clinical psychology in introductory psychology textbooks. *Teaching of Psychology, 18*, 12-16.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43*, 385-388.

Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector.* New York: McGraw-Hill.

McConnell, J. V., & Philipchalk, R. P. (1992). *Understanding human behavior* (7th ed.). Orlando, FL: Holt, Rinehart, and Winston.

Meyers, D. G. (1992). *Psychology* (3rd ed.). New York: Worth.

Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation -- A technical memorandum* (OTA-TM-H-15). Washington, DC: U. S. Government Printing Office.

Ornstein, R., & Carstensen, L. (1991). *Psychology: The study of human experience* (3rd ed.). New York: Harcourt Brace Jovanovich.

Patrick, C. J., & Iacono, W. G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology, 76*, 229-238.

Paul, D. B., & Blumenthal, A. L. (1989). On the trail of Little Albert. *The Psychological Record, 19*, 547-553.

Peterson, C. (1991). *Introduction to psychology.* New York: Harper Collins.

Pettijohn, T. F. (1992). *Psychology: A concise introduction* (3rd ed.). Guilford, CT: Dushkin Publishing.

Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology, 15*, 344-358.

Raskin, D. C. (1976). *Reliability of chart interpretation and sources of errors in polygraph examinations* (Report No. 76-3, Contract No. 75-NI-99-0001). Washington, D. C., National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U. S. Department of Justice.

Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional and legal issues surrounding the application and acceptance of polygraph evidence. *Utah Law Review, 1986*, 29-74.

Raskin, D. C. (1987). Methodological issues in estimating polygraph accuracy in field applications. *Canadian Journal of Behavioral Science, 19*, 389-404.

Raskin, D. C. (1989). Polygraph techniques for the detection of deception. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence.* New York: Springer. (247-296).

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15*, 126-136.

Raskin, D. C., Honts, C. R., & Kircher, J. C. (in press). The scientific status of research on polygraph techniques: The case for polygraph tests. Chapter to appear in D. L. Faigman, D. Kaye, M. J. Saks, & J. Sanders (Eds.), *The West companion to scientific evidence.*

Raskin, D. C., Kircher, J. C., Honts, C. R., & Horowitz, S. W. (1988). *A study of the validity of polygraph examinations in criminal investigation.* (Grant No. 85-IJ-CX-0040, National Institute of Justice). Salt Lake City: University of Utah, Department of Psychology.

Roediger, H. L., III, Capaldi, E. D., Paris, S. G., & Polivy, J. (1991). *Psychology* (3rd ed.). New York: Harper Collins.

Roig, M., Icochea, H., & Cuzzucoli, A. (1991). Coverage of parapsychology in introductory psychology textbooks. *Teaching of Psychology, 18,* 157-160.

Rubin, Z., Peplau, L. A., & Salovey, P. (1993). *Psychology.* Boston: Houghton Mifflin.

Santrock, J. W. (1988). *Psychology: The science of mind and behavior* (2nd ed.). Dubuque, IA: Wm. C. Brown.

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

Shaver, K. G., & Tarpy, R. M. (1993). *Psychology.* New York: Macmillan.

Shepard, R. N. (1983). "Idealized" figures in textbooks versus psychology as an empirical science. *American Psychologist, 38,* 855.

Slowik, S. M., & Buckley, J. P. (1975). Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure, and GSR recordings. *Journal of Police Science and Administration, 3,* 305-309.

Smith, R. E. (1993). *Psychology.* St. Paul, MN: West.

Soper, B., & Rosenthal, G. (1988). The number of neurons in the brain: How we report what we do not know. *Teaching of Psychology, 15,* 153-156.

Steller, M., Haenert, P., & Eiselt, W. (1987). Extroversion and the detection of information. *Journal of Research in Personality, 21,* 334-342.

Stern, R. M., Breen J. P., Watanabe, T., & Perry, B. S. (1981). Effects of feedback of physiological information on responses to innocent associations and guilty knowledge. *Journal of Applied Psychology, 66,* 677-681.

Suedfeld, P., & Coren, S. (1989). Perceptual isolation, sensory deprivation, and rest: Moving introductory psychology texts out of the 1950s. *Canadian Psychology, 30,* 17-29.

Szucko, J. J., & Kleinmuntz, B. (1981). Statistical versus clinical lie detection. *American Psychologist, 36,* 488-496.

Wade, C., & Tavris, C. (1993). *Psychology* (3rd ed.). New York: Harper Collins.

Waid, W. M., Orne, E. C., Cook, M. R., & Orne, M. T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of deception. *Journal of Applied Psychology, 63,* 728-733.

Weiten, W. (1992). *Psychology: Themes and variations* (2nd ed.). Belmont, CA: Brooks/Cole.

Weiten, W. (1994). *Psychology: Themes and variations* (2nd ed., briefer version). Belmont, CA: Brooks/Cole.

Wicklander, D. E., & Hunter, F. L. (1975). The influence of auxiliary sources of information in polygraph diagnoses. *Journal of Police Science and Administration, 3*, 405-409.

Winton, W. M. (1987). Do introductory textbooks present the Yerkes-Dodson Law correctly? *American Psychologist, 42*, 202-203.

Worchel, S., & Shebilske, W. (1992). *Psychology: Principles and applications* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Wood, E. R. G., & Wood, S. E. (1993). *The world of psychology.* Needham Heights, MA: Allyn and Bacon.

Wortman, C., Loftus, E., & Marshall, M. (1992). *Psychology* (4th ed.). New York: McGraw-Hill.

Zimbardo, P. G., & Weber, A. L. (1994). *Psychology.* New York: Harper Collins.

Appendix A: Factual Errors and Misdescriptions in the Text Excerpts.	
Text	*Errors and Misdescriptions* **Correct Information**
Atkinson et al.	*States that a relaxed baseline is taken for comparison to later responses.* **No polygraph tests do this.** *Person may be able to beat the test by causing reactions during the neutral questions.* **This would have no impact on the evaluation of a polygraph.** *The recording shown in the figure is referred to as a heart rate recording.* **It is a relative blood pressure recording.** *Persons who are less socialized may be less aroused and harder to detect.* **All of the empirical evidence suggests that this is not the case.**
Baron	*Control questions are described as name, place of birth, where someone works.* **These are neutral not control questions.**
Bernstein et al.	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.** *For the polygraph to be effective, the person being tested must believe that the machine is infallible in its ability to detect lies.* **No one who does research in this area states this position. There is no empirical research to support it and a great deal of research to refute it.**
Bootzin et al.	*The text suggests that you can beat the test with countermeasures to neutral questions.* **Countermeasures against neutral questions would have no effect.**
Carlson	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.** *The text describes a directed lie control test, but calls it a control question test.*

	The text states that the chance of a false positive error on a 3 key 5 item GKT is 8/1000. **The correct value is 1/125, i.e., 1/5 X 1/5 X 1/5, if the items are truly independent.**
Crooks and Stein	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Darley et al.	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Doyle	No errors.
Dworetzky	*States that there are separate channels for respiratory rate, heart rate, blood pressure and GSR.* **Heart rate is not measured unless it is derived from the blood pressure recording.** *The text indicates that subjects will be monitored while giving narrative answers to questions like "Where were you last night?"* **In actual tests all questions are answered "Yes" or "No".** *The date for Marston supporting the polygraph was given as 1932.* **Marston testified in U.S. v. Frye in 1923.** *The text states that most polygraph tests are given by employers and gives an example of a grocery store employee taking a screening test.* **Such tests were outlawed by the U. S. Congress in 1988.** *The text states that Honts, Hodes, and Raskin (1985) showed that it was "quite easy" to beat the polygraph by creating responses to truthful questions.* **Honts et al., instructed their subjects to increase their response to deceptively answered control questions in the context of a training session where subjects were fully informed about the nature and scoring of the test. With this intensive training only about half of the subjects could beat the test. Without training, none of the subjects were able to beat the test.** *States that Floyd Faye failed two polygraph tests.* **Faye failed one polygraph, the other was so distorted by Floyd's deliberate movements that it was not able to be scored.**
Feldman	*Polygraph measures irregularity in breathing pattern and increases in heart rate.* **The polygraph measures respiration, but irregularities are not scored. Heart rate is not scored.** *Biofeedback can be used to defeat the polygraph.*

	There is no evidence in the studies cited to support this assertion. Moreover, there are no credible data to support it in any source. *States that Honts, Hodes, and Raskin (1985) indicates that pressing on a tack in the shoe will allow people to beat the test.* **No such manipulation was included in Honts et al. (1985).**
Gleitman	No errors.
Huffman et al.	*States that polygraph tests can be fooled by people who take tranquilizers, who have consumed high levels of alcohol, or who are psychopathic.* **No cites are provided to support these statements. The empirical literature does not support any of them. The data on psychopaths is particularly clear. They have no special ability to fool the polygraph.**
Kalat	No errors.
Laird and Thompson	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.** *Faye's story about teaching other inmates how to beat the test is presented as fact.* **In reality, Faye's story is hearsay of hearsay from convicted felons. There is no evidence that anyone even took a polygraph and talked to Faye about it. This clearly is not scientific evidence.**
Lefton	*Habitual liars show little or no autonomic reactivity when they lie.* **The cite provided does not address this issue empirically. The literature indicates that psychopaths are just as detectable as normals.**
Meyers	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Peterson	No errors.
Pettijohn	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Roediger et al.	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Rubin et al.	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Santrock	*Polygraph relies on heart rate.* **Heart rate is not used.** *Drugs and biofeedback can be used to beat the test.* **The Waid et al. study failed to replicate. ALL other drug studies have failed to find effects. The Corcoran et al. study addresses the guilty knowledge test which is not in use in the field. There is no evidence to suggest that biofeedback can be used as a**

	countermeasure against actual field techniques. *Honts et al. is reported as showing that 80 percent of physical countermeasures could be detected by examiners.* **Honts et al. actually reported that most physical countermeasures could NOT be detected.**
Shaver and Tarpy	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.**
Smith	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.** *The recording shown in the figure shows one tracing as Pulse Rate Averaging.* **PRA is not used in polygraph. The tracing shown is a relative blood pressure tracing.** *Faye's story about teaching other inmates how to beat the test is presented as fact.* **In reality, Faye's story is hearsay of hearsay from convicted felons. There is no evidence that anyone even took a polygraph and talked to Faye about it. This clearly is not scientific evidence.**
Wade and Tavris	*Increased heart rate used as an indicator.* **Heart rate not used.** *People can learn to beat the machine by tensing muscles or thinking about an exciting experience during neutral questions.* **This would have no impact on the evaluation of a typical field polygraph.** *States that there are problems with reliability.* **The literature shows that the reliability of numerical scoring of the Control Question Test is very high, interrater reliabilities are almost always reported to be above 0.90.**
Weiten	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.** *States that critical questions are compared to nonthreatening questions.* **Critical questions are compared to Control Questions that are probable lies.** *Kleinmuntz and Szucko (1984) is described as an experiment.* **Kleinmuntz and Szucko is an archival field study.**
Weiten (briefer	*Heart rate is described as a dependent measure.*

version)	**Heart rate is not used as a dependent measure in the field.** *Text indicates that test questions have narrative answers.* **In the field all questions must be answered with either a "Yes" or "No".** *Kleinmuntz and Szucko (1984) is described as an experiment.* **Kleinmuntz and Szucko is an archival field study.**
Wood and Wood	*Text suggests that there is no pretest, that subjects are unaware of the wording of questions, and that subjects give narrative answers.* **There is a lengthy pretest were the test is explained and all of the questions are reviewed. Subjects must give "Yes" or "No" answers.** *The nature of the answer to the control questions is unimportant.* **The subject is maneuvered into answering the control questions with a deceptive response. The test is based on differential reactivity between relevant and control questions.** *Heart rate is listed as a dependent measure.* **Heart rate is not used in the evaluation of polygraph tests.** *Habitual liars are more likely to pass.* **There is no empirical evidence that this is true.** *Waid et al. (1981) cited as source of mental countermeasure study (counting backward by sevens).* **This study was actually Honts (1986).** *Lykken's (1981) popular book is cited as the source for drug and countermeasure studies.* **Although some countermeasure studies are discussed in Lykken (1981) no original data by Lykken are presented.** *Countermeasures during neutral questions are described as effective.* **Countermeasures during neutral questions would have no effect.**
Worchel & Shebilske	*Heart rate is described as a dependent measure.* **Heart rate is not used as a dependent measure in the field.** *Operators avoid asking did-you-do-it questions.* **Relevant questions are did-you-do-it questions. They are asked in virtually all tests.**

	The guilty knowledge test described as if it is the most common in the field. **The GKT is rarely used in the field.** *Faye's story about teaching other inmates how to beat the test is presented as fact.* **In reality, Faye's story is hearsay of hearsay from convicted felons. There is no evidence that anyone even took a polygraph and talked to Faye about it. This clearly is not scientific evidence.**
Wortman et al.	*Neutral questions are described as control questions.* *The control question test and the guilty knowledge test are mixed together in the general description of the techniques.*

Lexical Decision and the Detection of Concealed Information[1]

Lawrence Locker, Jr. and Marc E. Pratarelli

Abstract

The present study utilized the lexical decision task (LDT) to assess whether participants with concealed information regarding a subset of the word stimuli performed differently than those without. In Experiment 1, *Experimental* test subjects believed they were concealing special knowledge of certain word items from the Experimenter. In Experiment 2, *Informed* test subjects did not attempt to conceal their knowledge of the word items because they had received instructions from the Experimenter to study the items overnight. The Informed group responded more like Controls who possessed no special knowledge of the stimuli in the LDT. However, the Experimentals, having to conceal their knowledge of certain word stimuli, responded slower than either of the other two groups on the test items as well as on neutral word items and nonwords. The reaction time effects for different classes of words and the nonwords were constant across all three groups, an indication that prelexical processes are not affected. However, deceit appears to negatively impact word recognition speed by slowing all responses during postlexical processing.

This study examines the efficacy of using a well known speeded reaction time paradigm, the lexical decision task (LDT), in the detection of deceit of individuals who possess *concealed information.* For this study, concealed information refers to privileged information that one group possesses and another group does not. The study represents a departure from previous research in that it does not make use of the traditional methods and paradigms of lie detection, such as the Guilty Knowledge Test or the Control Question Test coupled with the polygraph or Interrogative Polygraphy (Barland & Raskin, 1975; Farwell & Donchin, 1991; Lykken, 1959; Reid & Inbau, 1977). Instead, the present study involves the detection of a special set of acquired knowledge consisting of certain words that were afforded special status before the test in certain subjects. Thus, the study attempted to ascertain whether subjects who had acquired special knowledge would manifest a bias in their patterns of responding in the LDT. In addition, because emerging technologies like Interactive Polygraphy have methodological constraints which require timelocking to the onset of a stimulus, another purpose of the study was to explore new behavioral approaches needed to accommodate these technological concerns. Hence, the present report is the result of a pilot series aimed at developing task demands that would reveal deception and conform to the methodological constraints of collecting event-related brain potentials (ERP).

[1]This article is reprinted with permission from *The Journal of Credibility Assessment and Witness Psychology* taken from Volume 1(1) pages 33 – 43. Please refer to http://truth.boisestate.edu for copies for reproduction under the original copyright that follows:

The conventional approach to lie detection utilizes the polygraph in conjunction with behavioral paradigms like the Guilty Knowledge Test (GKT) or the Control Question Test (CQT). The polygraph measures autonomic responses to questions involving critical events, e.g., of a crime, and other relevant and irrelevant questions in the hopes that the guilty knowledge items will elicit a stronger response from the guilty individual. Moreover, the physiological responses should also reject innocent individuals because they would not be physiologically aroused to the critical questions, by virtue of their not having specific knowledge of the event. The GKT involves a number of key items about the crime and a number of irrelevant items. A person with the knowledge of the key items and little response to the irrelevant items should have different physiological reactivity than a person with no knowledge of the key terms. Alternatively, the CQT involves questions related to the crime, neutral questions, and control questions that concern basic honesty, e.g., "Did you ever use a fake ID?" It is assumed that a guilty person would be more reactive to the relevant crime questions than an innocent person; the latter would be more concerned with the control questions.

In a newer and somewhat different approach, interrogative polygraphy uses ERPs as a lie detection tool (Farwell & Donchin, 1991; Rosenfeld, Angell, Johnson & Qian, 1991), again coupled with a behavioral paradigm that usually contrasts relevant and irrelevant information. This procedure calls for the analysis of highly visible brain waves elicited to isolated timelocked presentations of key words or pictures related to the crime. Stimulus items that subjects recognize as relevant should cause their brain to react with more activity than irrelevant items, even though a verbal or manual response may not be required. While the interactive polygraphy methods or the conventional polygraph approaches have met with varying degrees of success and acceptance, all of them continue to struggle with the difficult task of measuring the illusive psychological constructs of deceit and guilt. Moreover, none can claim to eliminate the number of false positives that continue to pose a problem in the practical application of these technologies. Presently, the aim of the study reported here is to examine the efficacy of another approach to measuring performance differences partially or wholly attributable to deceit or guilt.

The logic underlying the two experiments in the present study is based upon the distinction between automatic and controlled information processing. Automatic processes are not consciously controlled, i.e., they are normally outside a subject's awareness, and they are triggered very quickly at the mere presentation of the physical characteristics of the stimulus (Kihlstrom, 1987). In contrast, controlled processes are normally accompanied by awareness or access to consciousness. Controlled processes also require more processing time than automatic processes. The important theoretical issue for the present study is that the more cognitive demands/processes recruited during the task in a given individual, the more processing time they will need.

Concealing guilt or deceiving another individual involves a conscious and controlled effort to avoid acknowledging that one possesses a critical and relevant class of information. The present study attempts to circumvent the controlled cognitive processes and examines whether the automatic processes beyond a person's control are differentially affected during the processing of concealed information. In the LDT, the individual sees randomly presented words or nonwords (e.g., GLUG), and responds as quickly as possible by indicating whether the target item was one or the other. Because a conscious effort is focused on the word-nonword decision required by the task demands of the LDT, participants may not be able to dedicate enough attentional resources to their reactions to different classes of words nested within the word condition. Therefore, an individual hoping to use their concealed information to their advantage, may incidentally, generate side effects quite beyond their control. In principle, this should produce a behavioral marker of the possession of concealed relevant information.

Experiment 1

The focus of Experiment 1 was to assess the performance differences between individuals who have concealed information regarding certain items that will appear in their tests, and a set of control subjects who were not

offered information. Therefore, the contrast was between those who concealed their knowledge, and attempted to use it to their advantage, and those who concealed nothing. The experiment was not only aimed at detecting which group had the relevant information, but also to evaluate which behavioral effect was most sensitive for differentiating those who concealed their knowledge of that information.

Method

Participants

Forty-two participants (ages 18-24 years) recruited from introductory psychology classes were offered generous amounts of extra course credit for participation in the study. All participants gave their signed consent to participate on two consecutive days; they were screened for neurological history, normal or corrected to normal vision, reading disabilities, handedness, years of education, and whether English was their first and predominant language. Participants were assigned to one of two groups, referred to as Controls and Experimentals, in a quasi-random fashion. Each was told that their name appeared only on the consent form which would be given to their Instructor in order to record their extra credit. Afterward, the forms were discarded. Critical to the integrity of the study were instructions which stated that all information related to their performance was completely confidential, and could never be traced back to their name. Confidentiality was achieved by coding each participant's data with a randomly generated alpha-numeric code.

Materials

Four paragraphs were chosen from randomly selected topics. These included a paragraph about the planet Mars, one about chemistry, another dealing with fish, and another about the boiling point of liquids. For the stimuli in the LDT, 150 items/trials composed of 75 words and 75 pronounceable nonwords were generated in the following manner. Twenty-five words which appeared two or more times in the paragraphs, were selected as one category of word items. These words were designated the Relevant items. These words were matched to 50 additional words in the stimulus set on the basis or word length and word frequency of occurrence using the norms of Francis and Kucera (1982). Twenty-five of these latter 50 word items were designated Irrelevant items, i.e., they did not appear in the paragraphs and were not distributed in study lists. The last 25 word items were designated DOD-words because they comprised the concealed information. The DOD words also did not appear in the paragraphs, but certain participants believed these words would appear on the recognition/recall task given on the second day. The balance of the stimulus set consisted of 75 nonwords matched with the words for length and bigram frequency. All four stimulus categories were randomly ordered in the stimulus list. The list was presented in blocks of 50 trials.

Procedures

Each participant was informed that the experiment involved listening to one experimenter reading the four paragraphs to them on the first day, and that a recognition and recall task on a personal computer would be conducted on the subsequent day. They were also instructed that they would be working with two different experimenters on each of the two days. On the first day, four paragraphs would be read to them, and their instructions were to listen attentively because certain items from the paragraphs would appear on the task (LDT) the following day. The more items they recognized and recalled on the second day, the better they performed, and the more extra credit bonus points they would earn. The participants were informed that they would receive two extra credit points just for showing up on both days. However, they could also receive up to three additional points, depending on their performance on the second day. All participants were told during their debriefing that they actually received all five points of extra credit regardless of the quality of their performance, along with instructions not to discuss the experiment. The Controls received no further instructions except to return the next day at the appointment time. Those in the Experimental group were offered a study list by a confederate posing as another participant who had finished his second day of the experiment in the adjoining room. The first experimenter made an excuse to leave the lab for about five minutes while this occurred. The confederate informed each participant in the Experimental group that he had also been given the list the day before and that it had been helpful in

obtaining the maximum amount of extra credit. He also asked them not to reveal to the experimenter that they had been given the list. The confederate laid the list on the desk and left the room. If the participant accepted and concealed the study list, that individual was included in the Experimental group. The list was offered to each Experimental participant in order to create a situation in which they believed they were genuinely concealing something from the second experimenter. In order to create a valid blind condition and to protect the anonymity of the participant, the confederate did not reveal to either of the experimenters which of them were in the Control group or the Experimental group. Furthermore, to maintain and ensure privacy and confidentiality, the randomly generated participant codes had either an E or a C added by the confederate to distinguish their data during analysis.

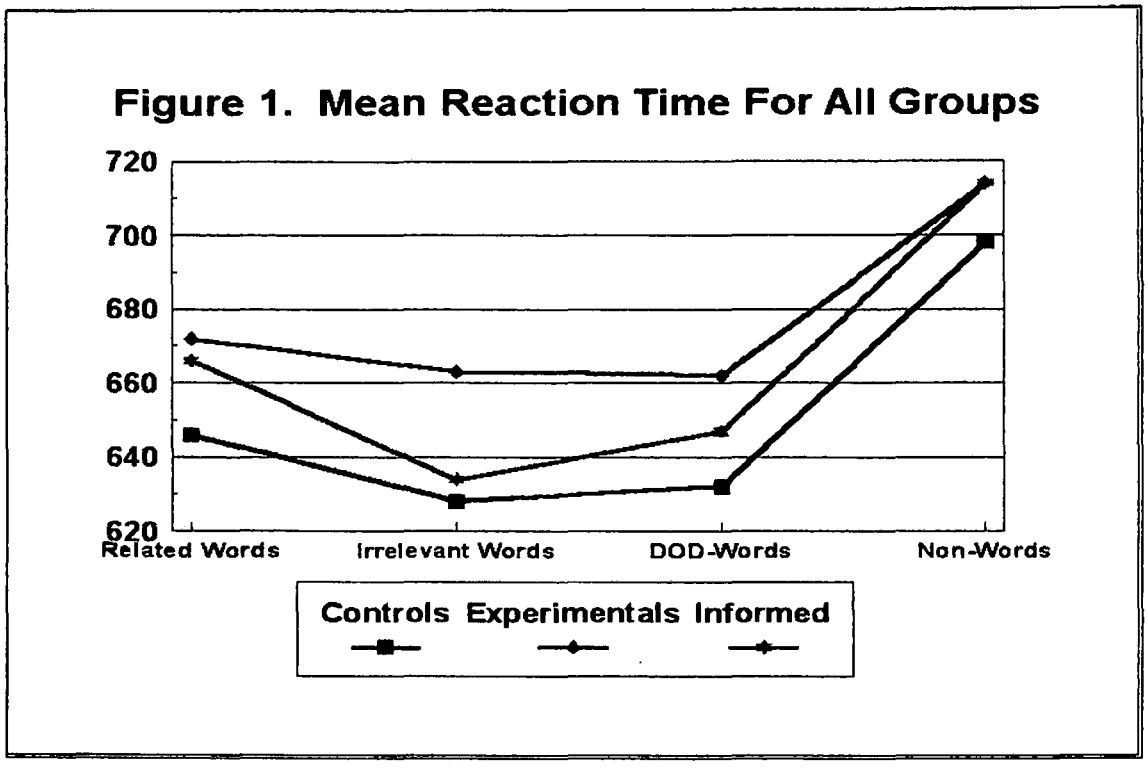
Results

A two (Group) by four (Word Condition) analysis of variance was used to analyze group and trial-type differences in reaction time in the LDT. Table 1 lists the means and standard errors for both groups in each condition. The trial type differences are also illustrated in Figure 1(red/diamonds and blue/squares). The analysis of variance produced a significant main effect of Group ($F[1,40] = 2.88$, $p < .045$), and a main effect of Word Condition ($F[3,120] = 32.52$, $p < .0005$). The analysis did not show a significant interaction effect between Group and Condition. A pooled average for all three Word Conditions versus the single Nonword Condition revealed that both groups responded faster to word targets than to nonword targets. Table 1 also lists the respective means and standard errors. These results are consistent with scores of previous studies using the LDT (cf., Forster & Taft, 1994; Pratarelli & McIntyre, 1994). No analysis was performed on this pooled data because the nonwords were nonconsequential to the focus of the study. However, because the respective means and standard errors do not overlap provides us with a reasonable assurance that both groups were performing and attending to the principal task of differentiating words and nonwords

Table 1. *Mean Reaction Times and Standard Errors for All Conditions in Experiment 1.*

Group		**Trial-type Conditions**			
		REL-W	**DOD-W**	**IREL-W**	**NON-W**
Controls	Mean	646.50	632.70	628.20	698.10
	S. Error	12.60	11.90	14.08	13.54
Experimental	Mean	672.40	662.50	662.90	734.30
	S. Error	17.64	16.18	11.16	20.56
			Pooled Word Conditions		**Nonwords**
Controls	Mean		635.80		698.10
	S. Error		12.81		13.54
Experimental	Mean		666.10		734.30
	S. Error		15.03		20.56

Legend: REL-W = Related Words; DOD-W = Detection of Deception Words; REL-W = Irrelevant Words; and NON-W = Nonwords. All means and standard errors are in milliseconds.



Discussion

Figure 1 reveals two important features of the data. First, Controls and Experimentals reacted comparably to word versus nonword targets. The word-nonword effect is important because it verifies that subjects were not directing their attention to the different categories nested in the word trials. During debriefing, none gave any indication that they were aware of different word categories when they were performing the LDT. The failure to recognize different word conditions provides a second assurance that both groups focused on the task demands of differentiating words from nonwords. Therefore, because both groups produced comparable word-nonword effects, we can reasonably ascertain that both were focused in similar ways, i.e., on making a rapid word-nonword decision.

The second and more important feature of the data, illustrated in Figure 1, was a standard delay across all four conditions for the Experimental group as compared to the Control group. That is, the Experimental participants reacted slower overall across all four trial types. Because the delay occurred in all trial-type conditions, it can be attributed to an additive postlexical effect rather that one which differentially affected some conditions and not others. Clearly, participants in the Experimental group responded differently to all conditions, not just to the DOD-words which they presumed they were studying without the knowledge of the experimenters. Moreover, because there were no differential effects between groups in word-nonword processing, the results suggest that the prelexical processes were not affected by the treatment variable, i.e., that the Experimental group had concealed information about the DOD items. Therefore, previous exposure to a study list, however covert, did not alter the early stages of word recognition. Such effects, if they had occurred, would have to be considered prelexical; that is, they took place prior to recognizing the item. Effects which alter processing by delaying all responses (words and nonwords) must occur postlexically. These findings can be reconciled using Posner and Snyder's (1975) two-process model in which the first stage is characterized by automatic stimulus processing, while the second stage is affected by the later conscious controlled processes. Presently, because no effects were

observed which could be attributed to the automatic (prelexical) stage, the standard delay for the Experimental group must be attributed to a late conscious controlled factor (postlexical). Such a factor would require some attentional resources.

The primary difference between the two groups was in terms of the awareness each had about the possibility that certain words would appear during the LDT. The Experimental group in particular believed that because they had concealed information, they would benefit from having studied the list acquired from the confederate. The Controls had no such belief because there was no concealed information. However, in order to benefit from the concealed information, the Experimental group had to focus more attention and effort on their accuracy at the expense of speed. The conventional framework for the speed-accuracy tradeoff is that depending on specific task instructions, task demands, or the influence of particular treatment variables, participants may sacrifice one at the expense of the other (Pachella, 1974). An accuracy analysis was not possible because both groups performed at ceiling levels. However, because the stimulus list was randomized, participants could not predict when a DOD word would appear. Therefore, participants in the Experimental group would have to commit more attentional resources and effort on their accuracy at the expense of reaction time. More attention would have to be diverted from the word-nonword discrimination task demands, and focused instead on the mere possibility that a word from their pool of concealed information would appear. The delay in mean reaction times for all four trial-types for the Experimental participants reflected their attention to accuracy, which in turn was the result of their attempts to utilize the concealed information from the acquired list.

Similar postlexical effects have been observed with social loafing (Pratarelli & McIntyre, 1994), heightened vigilance (Harkins & Petty, 1982), and when having to solve mazes (Jackson & Williams, 1985). Generally, altering any of the task demands which must compete for the limited attentional resources will impact reaction time postlexically. Another issue that should be considered is the self-efficacy of the Experimental participants (Bandura, 1988). It is reasonable to suppose that the Experimental participants assumed they would perform better due to their concealed knowledge. They focused more on the use of this knowledge instead of entirely devoting their attention to the task. In contrast, the Control participants assumed that their success was entirely dependent on the principal task demands, and their memory of the paragraphs heard the previous day. Therefore, Controls only took into account the instructions of the experimenters. Using this logic, Experimental participants, relying more on their belief that the concealed information would improve their performance, made inaccurate assumptions about the task. The effect of such cognitions was to draw on the limited attentional resources which, as seen in the better performance of the Controls, ought to be focused on the speeded word-nonword decision. The result is that Experimental participants were faced with having to implicitly divide their attention, while the Controls focused on a singular task.

In terms of alternate explanations for these results, it is generally accepted that concealing special knowledge creates a state of arousal in many individuals; this continues to be the underlying premise for the use of the polygraph. It is possible that a higher state of arousal could impact processing time and account for the differences observed between the two groups. However, the weakness and questioned reliability of the polygraph method, which hinges on higher arousal levels to detect deception, suggests arousal might have only minimal impact.

Another possibility is that participants could have been trying to overanalyze the task on the second day. On the first day, those who left with the list and a brief description of the next day's task, left with the crude understanding that somehow the list would improve their ability to recall those items. Unbeknownst to them, however, the LDT is not a recall task. Therefore, participants were faced with an unfamiliar task which was not consistent with their expectations of how the information would be useful to them. Yet another consideration was that the Experimental participants may have processed the words on at least two levels as compared to the Controls. These levels involve (1) the obligatory recognition of the stimulus, and (2) the later processing component involving the

realization that it was or was not a word from their list. However, if this had occurred, the DOD items would have been processed differently than the other word conditions, but they were not. Indeed, the most reasonable explanation is that the metaknowledge of having concealed information slowed all their responses, because it could never be predicted that a known (DOD) word would appear on any given trial. However, a question remains whether simply having the privileged access to information, rather than the guilt related to that knowledge, is the driving force behind the reaction time delay. Therefore, we determined that further study was needed to focus solely on the effects of privileged knowledge without any guilt, concealment or deception.

Experiment 2

The focus of the second experiment was to assess differences in a new group of participants given the privileged knowledge without any attempt to hide that knowledge. The purpose of this experiment, therefore, was to assess whether poorer performance in the Experimental group earlier was actually due to mere access to privileged information, or to the fact that they believed they had concealed information. If indeed the slower response times for the Experimental group were due to their increased attention on the material they had studied, then anyone having similar opportunities to study in advance ought to manifest comparable delays.

Method

Participants

Twenty-one additional students (ages 18-24 years) were recruited for the new treatment condition. These were referred to as Informed participants. Because we decided to compare them to the existing Control group, participants could not be randomly assigned. The new participants received the same instructions, screening, coding, and extra credit as the participants in Experiment 1.

Stimuli and Procedures

The same stimuli in the same randomized order were used in Experiment 2. The procedures differed in that the Experimenter gave all the participants the study list after reading the four paragraphs to them. The Experimenter added that the items on the study list would appear in the task the next day, and they would have a better chance of receiving all the extra credit points if they studied the list items. No further instructions were given except to return the next day at their appointment time to complete the computer task.

Results

A two (Group) by four (Word Condition) analysis of variance was utilized to analyze group and trial-type differences in reaction time in the LDT. Table 2 lists the means and standard errors for the Informed group along with the means and standard errors from the Control group in Experiment 1. An analysis of variance did not reveal a significant main effect of Group, but only a main effect of Word Condition ($F[3,120] = 34.11$, $p < .0005$). The analysis also showed no significant effect in the Group by Word Condition interaction. A pooled average for all three word conditions/trials versus the nonword trials again revealed that both groups responded faster to word targets than to nonword targets. Table 2 lists the respective means and standard errors compared to the Controls. As in Experiment 1, these results provide a reasonable assurance that the focus of the participants' attention was on differentiating words from nonwords. The data for the Informed group were subsequently plotted against those from Experiment 1 (Figure 1, green/stars)

Discussion

There are two notable features in the present experiment. First both Control and Informed participants responded comparably to word and nonword items. This effect is important in that it directs the participants' attention away from the word categories. The second notable feature of the new data is that the Informed participants did not respond significantly different than the Control group in Experiment 1. However, the data from Table 2, and as illustrated in Figure 1, reveal that there appears to be a small standard delay across all four Word Conditions. Although not statistically significant, this effect bears some resemblance to the pattern produced by the Experimental group earlier. The delay for the Informed group is different in that it more closely approximates the performance of the

Table 2. *Mean Reaction Times and Standard Errors for All Conditions in Experiment 2.*

Group		**Trial-type Conditions**			
		REL-W	**DOD-W**	**IREL-W**	**NON-W**
Controls	Mean	646.50	632.70	628.20	698.10
	S. Error	12.60	11.90	14.08	13.54
Informed	Mean	665.70	646.70	633.80	713.90
	S. Error	17.56	14.48	15.21	14.64
		Pooled Word Conditions			**Nonwords**
Controls	Mean		635.80		698.10
	S. Error		12.81		13.54
Informed	Mean		649.10		713.90
	S. Error		15.49		14.64

Legend: REL-W = Related Words; DOD-W = Detection of Deception Words; IREL-W = Irrelevant Words; and NON-W = Nonwords. All means and standard errors are in milliseconds.

Controls than the Experimentals. A similar analysis of variance comparing the Informed group with the Experimental group was also not significant. Despite the small similarity, the difference in the standard delay between Informed and Experimental groups may be attributed to the manner in which their special knowledge was framed, i.e., nonconcealment versus concealed.Both Experimental and Informed participants received the special knowledge for some of the stimulus items with the instructions that it would "make it easier to get" extra credit points. Because the Informed participants did not have to focus any effort on concealing their special knowledge from the Experimenter, one possible explanation for their slightly faster performance, when compared to Experimentals, is that the guilt and deceit related to the knowledge somehow usurped some of the attentional resources needed for responding quickly. A mechanism by which this happens remains to be explored. Using similar reasoning, the Informed participants focused more attention on the task demands because there was no concealment associated with their knowledge of the list. Although not significant, the small but uniform differences between the Informed and Control participants might be explained in the same way by saying that merely having special knowledge and the idea that the knowledge would improve performance, slowed their response times slightly, but not significantly. The Experimental group possessed the same idea but were further impacted by the need to conceal their having acquired the list. Using this rational, the conscious cognitive process of concealing privileged information draws on limited attentional resources to some degree.

Given the slight but nonsignificant delay between the Controls and the Informed group shown in Figure 1, we conducted a *post-hoc* power analysis to estimate how many participants would have been required to achieve a significant effect with that degree of variability. The results suggest that approximately 150 participants would have been necessary to achieve significance. This is important because the interpretation of the present results would have to be altered somewhat to accommodate the mere effect of being Informed, versus being informed but having to conceal the knowledge. It is also important inasmuch as ERP studies normally have group sizes ranging from 10 to 30 participants; more than 30 is generally not feasible. However, many ERP studies have demonstrated that ERP effects can be very large even in the absence of significant behavioral effects.

On the whole, participants in the Informed and Experimental groups performed progressively slower as a function of the amount of resources each group may have dedicated to the process of consciously recognizing individual words during the LDT. Because nonwords were also affected, but were

never candidates from the study list, the effect on the attentional system was to draw resources away from the task which in turn slowed the response to all target items. In terms of final conclusions, concealed information may alter certain cognitive processes which may differentiate performance on certain attention demanding tasks. Further study is required to examine the effects of other possible situational variables as well as individual differences. Moreover, the timebase of the effects of concealing one's knowledge of information can be further understood by adapting the LDT or similar paradigms with newer technologies like ERPs and interrogative polygraphy.

Author Notes

We would like to thank John Locker and Alvin Sunni for their performance as confederates, and Mary Devitt for her editorial suggestions and statistical support. We also wish to thank all of our student participants. Address correspondence to: Marc E. Pratarelli, Ph.D., 215 North Murray, Department of Psychology, Oklahoma State University, Stillwater, OK 74078.

References

Bandura, A. (1988) Self efficacy: toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Barland, G. H., & Raskin, D. C. (1975) An evaluation of field techniques in detection of deception. Psychophysiology, 12, 321-330.

Farwell, L. A., & Donchin, E. (1991) The truth will out: interrogative polygraphy ("Lie Detection") with event-related brain potentials. Psychophysiology, 28, 531-547.

Forster, K. I. & Taft, M. (1994) Bodies, antibodies, and neighborhood density effects in masked form priming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 455-464.

Francis, W. N., & Kucera, H. (1982) Frequency analysis of English usage: Lexicon and grammar. Boston: Houghton-Mifflin.

Harkins, S. G., & Petty, R. (1982) Effects of task difficulty and task uniqueness on social loafing. Journal of Personality and Social Psychology, 43, 1214-1229.

Jackson, J. M., & Williams, K. D. (1985) Equity in effort: an explanation of the social loafing effect. Journal of Personality and Social Psychology, 49, 937-942.

Kihlstrom, J. F. (1987) The cognitive unconscious. Science, 237, 1445-1452.

Lykken, D. T. (1959) The GSR in the detection of guilty knowledge. Journal of Applied Psychology, 43, 385-388.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B.H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition (pp.41-82). New York: Wiley.

Posner, M. I., & Snyder, C. R. R. (1975) Attention and cognitive control. In R.L. Solso (Ed.), Information Processing in Cognition: The Loyola Symposium. Hillsdale, N.J.: Erlbaum.

Pratarelli, M. E., & McIntyre, J. A. (1994) Effects of social loafing on word recognition. Perceptual and Motor Skills, 78, 455-464.

Reid, J. E., & Inbau, F. E. (1977) Truth and Deception: The Polygraph Technique. Baltimore: Williams & Wilkins.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. (1991) An ERP-based, control-question lie detector analog: algorithms for discrimination effects within individuals' average waveforms. Psychophysiology, 28, 319-335.

The Effect of Attorneys' Nonverbal Communication on Perceived Credibility[1]

Patricia Rockwell[2] and Amy Ebesu Hubbard

Abstract

This study examined how attorneys' nonverbal communication relates to perceptions of credibility. Types of nonverbal communication considered included vocalic (fluency, pausing, variety), kinesic (facial expressiveness, number of illustrative gestures and dynamic quality of gestures), and physical appearance (grooming, age, facial hair, height, weight, and attractiveness). Four dimensions of credibility were considered--competence, trustworthiness, friendliness, and dynamism. Attorneys with greater facial expression and greater pitch variety were perceived as less competent. Those with greater facial expression, pitch variety, and tempo variety were perceived as less trustworthy. However, attorneys with greater pitch and tempo variety were perceived as more friendly. These findings highlight the dilemma that attorneys experience when attempting to improve their perceived credibility. They must be aware that credibility has multiple dimensions and that behaviors that enhance one aspect may detract from other aspects.

Introduction

Although a courtroom trial is typically thought to be a verbal exercise in debate and argumentation, nonverbal cues also play an important role (Barge, Schleuter, & Pritchard, 1989). That is, jurors form impressions of lawyers' credibility based on what lawyers do as well as what they say (Goldberg, 1982; Haynes, 1984; LeVan, 1984).

An especially relevant point at which jurors form their first impression of an attorney occurs at the beginning of the trial during opening statements (Wrightsman, 1987). These initial impressions formed of the attorney may impact all judgments made by jury members throughout the trial in regards to the quality of the evidence presented and the arguments given.

Some researchers argue that nonverbal behaviors are more influential than verbal behaviors in determining first impressions (Berger & Calabrese, 1975; Sunnafrank 1986). Burgoon, Buller, and Woodall (1996) suggest that physical appearance is the most influential because it is the first cue one experiences when meeting someone for the first time. Indeed, jurors can observe attorneys' physical appearance before the attorneys begin to speak. In addition to physical appearance, vocal and kinesic (or body movement) cues also impact first impressions.

The present investigation examined the association between attorneys' nonverbal behaviors and credibility evaluations made of them by potential jurors during opening statements when it is most likely that jurors will form first impressions.

[1]This article is reprinted with permission from *The Journal of Credibility Assessment and Witness Psychology* taken from Volume 2(1), pages 1 - 13. Please refer to http://truth.boisestate.edu for copies for reproduction under the original copyright that follows:

[2]Please address all correspondence to: Patricia Rockwell, Department of Communication, University of Louisiana at Lafayette, Lafayette, LA, 70504. (318) 482-6090 (office); (318) 482-6104 (fax); EMAIL par2323@usl.edu.

Dimensions of Credibility

Credibility is a construct that consists of all the judgments made by receivers regarding a speaker's believability. Researchers propose that credibility is composed of various dimensions such as competence, trustworthiness, friendliness, dynamism, character, and sociability (McCroskey, Jensen, & Valencia, 1973; Miller & Hewgill, 1964; Smith, 1982). A variety of different combinations of these dimensions have been examined with most researchers noting that competence and character seem to be the most important to perceptions of credibility (e.g., Burgoon, Birk, & Pfau, 1990; McCroskey & Young, 1981).

Researchers have investigated the relationship between these six dimensions of credibility and various nonverbal behaviors during courtroom proceedings (Miller & Hewgill, 1964; Pearce & Brommel, 1972, Pearce & Conklin, 1971, Sereno & Hawkins, 1967). However, most of these studies have only investigated the perceived credibility of witnesses (Bradac, Hemphill, & Tardy, 1981; Erickson, Lind, Johnson, & O'Barr, 1978; Hemsley & Doob, 1978; Lindsay, Wells, & O'Connor, 1989; O'Barr, 1982) rather than those formed of attorneys. As lawyers are central figures in any trial, it is likely that their perceived credibility is as pertinent as witnesses' perceived credibility.

Furthermore, most studies that have examined the effects of various behaviors on perceived credibility have focused primarily on verbal behaviors or nonverbal behaviors closely related to verbal behavior such as fluency and pausing (Addington, 1971; O'Barr, 1982; Schweitzer, 1970) and have ignored other influential nonverbal factors such as gestures, facial expression, and physical appearance. The current study explored the relationships between a greater array of nonverbal behaviors such as vocalics, kinesics, and physical appearance with the perceived credibility of attorneys as they presented actual opening statements in the courtroom.

Vocalic Cues of Credibility

Most researchers have concentrated on only a few vocal cues and their relationship to credibility (Bradac, Hemphill, & Tardy, 1981; Erickson, Lind, Johnson, & O'Barr, 1978; Hosman & Wright, 1987; O'Barr, 1982). Vocalic behaviors generally investigated in credibility are those of fluency and pausing (Addington, 1971; O'Barr, 1982; Schweitzer, 1970). Fluency is generally defined as speech that is free of long pauses, hesitations, repetitions, and extraneous vocal sounds such as "um" or "hmm."

For example, Burgoon, Birk, and Pfau (1990) trained students to evaluate nonverbal communication and asked them to evaluate the credibility of classmates giving speeches. They found that greater fluency was associated with higher ratings on the competence, sociability, and composure dimensions of credibility. In another study, Barge, Schleuter, and Pritchard (1989) examined the influence of lawyers' vocal delivery on judgments of lawyer credibility made by receivers. These researchers found that a fluent style, in contrast to a nonfluent style, was associated with perceptions of lawyers as more competent and more dynamic, and that a nonfluent style, rather than a fluent style, was associated with perceiving lawyers as more friendly. Other research supports the notion that fluency is positively correlated with competence and dynamism (McCroskey & Mehrley, 1969; Miller & Hewgill, 1984; Sereno & Hawkins, 1967) and persuasiveness (Mehrabian & Williams, 1969). Thus, the following hypotheses are offered:

H1: Lawyers' vocal fluency will be positively correlated with participants' perceptions of lawyers' competence, trustworthiness, and dynamism.

H2: Lawyers' vocal fluency will be negatively correlated with participants' perceptions of lawyers' friendliness.

Fluency is a major vocal cue that has been shown to affect credibility, but other elements of the voice may also impact perceptions of credibility. For example, variations in tempo, pitch, and volume may contribute to vocal expressiveness., and vocal expressiveness may enhance assessments of lawyers' credibility. Burgoon and her colleagues (1990) found that greater pitch variety was positively correlated with perceptions of competence, sociability, and character. Scherer (1979) found a

significant positive correlation between pitch range and perceived influence. However, Barge and his colleagues (1989) found that, although varied pitch was associated with perceptions of speakers as more dynamic, it was also associated with perceiving speakers as less friendly and trustworthy. Taken together, these findings indicate that variations in tempo and pitch may have a positive impact on certain dimensions of credibility. Furthermore, variations in the vocal feature of volume may also impact perceptions of lawyer credibility, but it is uncertain exactly what that effect might be. Thus, the following hypotheses and research question are offered:

H3: Lawyers' vocal expressiveness (e.g., varied tempo and pitch) will be positively correlated with perceptions of lawyers' competence, trustworthiness, and dynamism.

H4: Lawyers' vocal expressiveness will be negatively correlated with perceptions of lawyers' friendliness.

RQ1: What effect will increased volume variety have on participants' perceptions of lawyers' competence, trustworthiness, dynamism, and friendliness?

Kinesic Cues of Credibility

Live face-to-face interaction in the courtroom yields a host of other nonverbal cues that provide additional information upon which jurors may form impressions of attorneys credibility. Indeed, nonverbal researchers emphasize the primacy of the visual channel in affecting perceived credibility (Ekman, Friesen, O'Sullivan, & Scherer, 1980; Krauss, Apple, Morency, Wenzel, & Winton, 1981; Zaidel & Mehrabian, 1969). Jurors may be just as likely (if not more likely) to rely on what they see as what they hear when judging attorneys' initial credibility.

Thus, in addition to vocal expressiveness, "kinesic expressiveness" may also be positvely associated with judgments of attorney credibility. Kinesics includes body movements such as gestures and facial expressions. Researchers generally distinguish between illustrative (or meaningful) gestures and adaptor (or nervous) gestures. Various kinesic behaviors have been shown to be relevant to assessments of credibility. For example, when evaluating public speakers, competency judgments were increased by the use of kinesic behaviors such as facial pleasantness (Burgoon et al., 1990). LaCrosse (1975) found that counselors who smiled, made eye contact, and gestured more often were perceived as more competent than counselors who did these things less often. Thus, the following hypothesis is offered:

H5: Kinesic expressiveness (e.g., facial expressiveness, illustrative gestures, use of dynamic gestures) will be positively correlated with perceptions of lawyers' competence, trustworthiness, friendliness, and dynamism.

Physical Appearance Cues of Credibility

In an actual courtroom, jurors have access to the entire spectrum of a lawyer's nonverbal behaviors. For example, jurors have access to physical appearance cues, such as grooming, age, facial hair, height, weight, and attractiveness. In a related vein, Burgoon and her colleagues (1989) speculated that clothing, for instance, may affect perceived client credibility in a legal setting. Several studies have indicated that adherence to conventional attire and good grooming can increase compliance (Crassweller, Gordon, & Tedford, 1972; Darley & Cooper, 1972). In one study, women with short hair, conservatively dressed, with earrings and a necklace were perceived as more competent than those without these features (Rosenberg, Kahn, & Tran, 1991). Bickman's (1971) classic study found that compliance was greater for experimenters dressed in suits and ties than for those wearing work clothes. Brownlow (1992) found that people with mature faces were more persuasive than people who had more childlike faces. Yet, few of these physical appearance cues have been considered in the context of jurors' perceptions of lawyer credibility. Thus, the following research question is posed:

RQ2: How do physical appearance cues relate to perceptions of lawyers'

competence, trustworthiness, friendliness, and dynamism?

In summary, this research sought to determine what effects, if any, the nonverbal communication behaviors of vocalics, kinesics, and physical appearance had on participants' perceptions of attorneys' credibility. Specifically, the effects of these behaviors on the credibility dimensions of competence, trustworthiness, friendliness, and dynamism were considered.

Method

Participants

Forty-eight undergraduate students (18 males and 30 females) were recruited from communication classes and offered extra credit for participation in this study. Participants were gathered in five groups that ranged in size from seven to 12 per group.

Stimulus Manipulation

Eight lawyers (three male and one female prosecuting attorneys and four male defense attorneys) presenting opening statements were recorded on videotape from "The Court Channel." This cable channel broadcasts live trials 24 hours a day. Five different trials were videotaped. Four of these trials were used in the primary analyses of this study and one was used for practice purposes. The four trials used in the experiment consisted of (a) Art Buchwald's suit against Paramount Studios concerning claims that the movie *Coming to America* was illegally based on his work, (b) an appeals case presented before the Arizona Supreme Court in which defense attorneys argued that hair evidence in the murder trial of Charles Treadway, Jr., had been improperly handled by police officers, (c), a Florida murder case against Virginia Larzelere of Daytona Beach, who was accused of plotting to kill her husband with the assistance of multiple male accomplices, and (d) the Rodney King police brutality case.

The eight original opening statements ranged in length from ten to twenty minutes. To keep length equivalent, one two-minute segment was extracted from approximately the middle of each opening statement. Thus, each master videotape consisted of two, two-minute segments from each trial--one of the prosecuting attorney and one of the defense attorney. Lawyers seen on the videotapes were presented in partial profile with only their upper torsos visible. In addition, each lawyer was either situated behind a podium or seated at a table.

Nonverbal Measures

Two researchers and two research assistants served as coders for this project. The researchers developed a coding form (Appendix A) designed to measure a variety of nonverbal behaviors that might be used in forming impressions of credibility. All variables were measured on a 9-point scale on which higher scores reflected more fluency, more vocal expressiveness, and more kinesic expressiveness. Higher scores for the physical appearance cues represented attorneys rated as better groomed, older, having less facial hair, taller, heavier, and more attractive.

The nonverbal behaviors included in the analysis along with their inter-item alpha reliabilities were (a) vocal fluency (.31), (b) vocal expressiveness--tempo variety, pitch variety, and volume variety (.70), (c) kinesic expressiveness--facial expressiveness, number of illustrators, and dynamic illustrators (.84), and (d) physical appearance cues--grooming, age, facial hair, height, weight, and attractiveness (.65). In addition, the four coders independently rated the nonverbal behaviors of the lawyers. Inter-rater alpha reliabilities for the each of the eight attorneys were .96, .93, .96, .94, .95, .90, .91, and .94, respectively.

Credibility Measures

Credibility was assessed by combining Miller and Hewgill's (1964) speaker credibility scale and Williams, Farmer, Lee, Cundie, Howell, and Rooker's (1975) scale of global perceptions of attorneys (Appendix B). The resulting credibility measure was originally factor analyzed by Barge and his colleagues (1989) and yielded four dimensions of credibility: competence, which was measured by bipolar anchors of knowledgeable/uninformed, precise/vague, accurate/inaccurate, certain/uncertain, expert/ignorant, trained/untrained, and competent/incompetent (interitem alpha reliability = .92); trustworthiness, which was measured by bipolar anchors of fair/unfair, telling the truth/not telling the truth, sincere/insincere, just/unjust, kind/cruel,

and admirable/contemptible (interitem alpha reliability = .79); dynamism, which was measured by bipolar anchors of aggressive/meek, bold/timid, energetic/tired, and extraverted/introverted (interitem alpha reliability = .86); and friendliness, which was measured by the bipolar anchors of warm/cold and open/defensive (interitem alpha reliability = .47). Items were rated on 9-point scales with higher scores representing more competence, more trustworthiness, more dynamism, and more friendliness.

Procedure

Upon arriving to the communication laboratory, participants were seated around a large table. A videocassette recorder was placed within clear view of all participants at the table. The research assistants explained the procedures to be followed and distributed the credibility questionnaires to each participant.

Next, participants were given the opportunity to view a practice videotape in order to familiarize themselves with the type of video they would see and the use of the credibility questionnaire. Following the practice tape, the researchers answered any procedural questions that participants had. Next, participants were shown one of the four master tapes. Presentation of tapes was counterbalanced to prevent order effects so that each group viewed the four tapes in a different order. Each tape consisted of two opening statements by two different lawyers. After participants viewed a particular lawyer's opening statement on the tape (in the order of prosecution then defense for each tape), the researchers stopped the tape so that participants could complete the questionnaire devoted to the evaluation of that particular lawyer. After participants viewed all four tapes and completed questionnaires on all eight lawyers' credibility, they were thanked for their efforts, debriefed regarding the nature of the study, and dismissed.

Results

Pearson product-moment correlational analyses were used to examine the associations between the various nonverbal cues and the four dimensions of credibility (i.e., competence, trustworthiness, dynamism, and friendliness). Although more sophisticated statistical tests could be used, correlational analysis appeared to represent an appropriate test for naturalistic data (Scherer, 1979). Also, because of the small sample size and the likelihood of reduced power, it may be appropriate to demand more stringent alphas from the resulting correlation coefficients.

In order to conduct the analyses, several scores were computed from both the nonverbal and the credibility measures. First, means scores were obtained for each of the four credibility dimensions using the scores for the individual credibility items. Correlations were then computed between the each of means for the nonverbal behaviors and the means for the four credibility dimensions (Table 1).

Hypothesis 1 and 2

Results indicated no support for Hypothesis 1 or 2. Fluency was not significantly correlated with competence, trustworthiness, friendliness, or dynamism.

Hypothesis 3 and 4 and RQ1

Contrary to expectations stated in H3 that greater pitch and tempo variety would be positively associated with perceptions of credibility, increases in pitch variety were negatively correlated with judgments of competence, $r = -.86$, $p < .01$, and trustworthiness, $r = -.96$, $p < .01$. Likewise, tempo variety was negatively correlated with judgments of trustworthiness, $r = -.80$, $p < .05$. Contrary to Hypothesis 4, which posited that greater tempo variety would be negatively associated with ratings on friendliness, tempo variety was found to be positively correlated with ratings of friendliness, $r = .71$, $p < .05$. In response to RQ1, increased volume variety was positively correlated with ratings of friendliness, $r = .73$, $p < .05$.

Hypothesis 5

No support was obtained for Hypothesis 5. Kinesic expressiveness was not positively correlated with any of the dimensions of credibility. Indeed, facial expressiveness was negatively correlated with competence and trustworthiness assessments. Lawyers who were more facially expressive were rated as significantly lower in competence, $r = -.75$, $p < .05$, and trustworthiness, $r = -.78$, $p < .05$.

Table 1. Correlation Coefficients

Variable	Credibility Dimension			
	Competent	Trust-worthy	Friendly	Dynamic
Vocal Fluency	.45	.12	.11	.16
Tempo Variety	-.65	-.80*	.71*	-.31
Pitch Variety	-.87**	-.97**	.66	-.54
Volume	-.32	-.54	.73	-.61
Facial Expressiveness	-.75*	-.78*	-.61	-.13
# of Gestures	-.30	-.34	.46	.30
# of Dynamic Gestures	-.35	-.42	.60	.17
Grooming	-.15	-.04	-.55	.17
Age	.56	.45	-.60	-.04
Height	.57	.49	-.40	-.03
Weight	.36	.31	-.31	.26
Facial Hair	-.21	-.20	.43	.44
Attractiveness	-.40	-.11	-.29	-.04

* $p < .05$ ** $p < .01$ all tests 2-tailed

Research Question 2

Analyses pertinent to research question 2 did not yield any significant correlations. None of the physical appearance cues (grooming, age, facial hair, height, weight, and attractiveness) significantly associated with assessments of competence, trustworthiness, friendliness, or dynamism.

Discussion

Although none of the hypotheses were fully supported, the present study's results suggest that some nonverbal behaviors may affect some dimensions of perceived credibility. Results indicated that four of the nonverbal behaviors investigated were significantly correlated with some of the dimensions of

credibility. Pitch variety and facial expressiveness showed a negative correlation with competency ratings. Pitch variety, tempo variety, and facial expressiveness were also negatively correlated with judgments of trustworthiness. Tempo variety and volume variety were positively associated with friendliness assessments. No significant correlations were obtained with dynamism.

That variety in pitch and tempo had such diametrically opposite effects on ratings of competence and trustworthiness does not support previous research findings. Also, if one views facial expressiveness as an indicator of variety, then it appears that, in general, jurors may view competent and trustworthy lawyers as those who exhibit little expressiveness in their nonverbal behavior. Possibly, the reason for these lower ratings on competency and trustworthiness is that attorneys' expressive behaviors may have been perceived as artificial or "actor-like." Perhaps expressiveness is not viewed as an appropriate behavior in courtroom settings and thus, lawyers who exhibit these behaviors may be seen as less competent. However, just the opposite appears to be true for attributions of friendliness, which were enhanced by increased variety in vocal behavior.

Limitations

Several limitations qualify the conclusions that can be drawn from this study. First, low interitem alpha reliability for friendliness make interpretation of its positive correlation with tempo and volume variety tentative. Second, because attorneys were seen standing behind a podium in a medium television shot, some nonverbal behaviors such as height and weight were difficult to determine. Third, fluency was not significantly correlated with credibility in this study, although many researchers have consistently found fluency a major contributing factor to credibility. Possibly, this finding was due to the low reliability scores for fluency or the fact that a lawyer experienced enough to argue a case noteworthy enough to be televised can be assumed to have a sufficiently fluent delivery.

Fourth, there was some lack of experimental control because naturalistic stimulus materials were used. However, Scherer (1979) suggests that there is a tradeoff between control and ecological validity when using naturalistic data. He argues that although lack of experimental control may present problems for analysis, it may ultimately lead to more generalizable data.

Fifth, there may have been insufficient variation between the nonverbal behaviors of the eight lawyers viewed. The sample of attorneys was small and probably not truly representative of the average lawyer arguing before a local jury.

Finally, although gender was not an issue considered in this study, one of the eight lawyers presented on the tapes was female. This may have influenced participants. Certainly, both the nonverbal and credibility scores obtained for the female attorney were not noticeably divergent from those of her male counterparts. Even so, this factor may be useful to consider in future studies.

Future Directions

Although previous research did not attempt to replicate natural courtroom interactions (Miller, Fontes, Boster, & Sunnafrank, 1983), future researchers can do so more easily. Past research has had participants silently read hypothetical transcripts of trials and then make evaluations (e.g., Hosman & Wright, 1987; Pearce & Brommel, 1972). More recently, Barge and his colleagues (1989) simulated a "real" courtroom with undergraduate students acting as jurors/participants (e.g., a judge giving directions, and the voices of law students on audiotape presenting opening statements which were derived from a transcript of an actual trial). The advent of videotape may offer even more realism (Scherer, 1982). With television networks such as "The Court Channel," now available, using videotapes of actual courtroom interactions to study attorneys is now feasible.

Future researchers might also consider the issue of when judgments or assessments of attorneys are made and test impressions at those critical junctures, instead of waiting until the end of the trial. Frequently, researchers have used client guilt--an end factor measure--as the sole measure of attorney credibility. However, other courtroom factors may impinge upon determinants of credibility besides ultimate determinations of client guilt. Further, credibility assessments

can change as a trial progresses--even during presentation of opening statements when, as we have argued, initial impressions of lawyer credibility are formed. Thus, the present investigation assessed credibility sequentially, with credibility being judged after the prosecuting attorney's opening statement, and then after the defense attorney's opening statement.

In addition, future researchers might investigate other vocalic, kinesic, and physical appearance cues and their association with attorney credibility in the courtroom. The current study is unique in that it incorporated vocal variety features into the analysis whereas most studies have generally examined a fairly small set of vocal behavior, such as pitch, tempo, and fluency.

Some researchers (e.g., Scherer, 1979) contend that visual cues are not as influential as their auditory counterparts in perceptions of credibility. However, although no significant correlations were found between credibility and physical appearance in the present study, it is possible that this may have been due to the lawyers investigated. Lawyers drawn from "The Court Channel" may represent a relatively homogeneous group, at least as far as outward appearance goes, in that these lawyers knew they would be televised nationally and thus were particularly careful in their dress and grooming.

Attorneys presenting opening statements represent a rich source of nonverbal behaviors with which to investigate jurors' perceptions of lawyers' credibility. The credibility impressions of lawyers that are formed and assessed throughout a trial may substantially impact courtroom interaction, and ultimately client guilt or innocence. Clearly, this is an important area that deserves more research attention. It is hoped that future researchers will benefit from this early tentative work on credibility judgments of more naturalistic courtroom behavior.

References

Addington, D. W. (1971). The effect of vocal variations on ratings of source credibility. *Speech Monographs, 38*, 242-247.

Barge, J. K., Schlueter, D. W., & Pritchard, A. (1989). The Effects of nonverbal communication and gender on impression formation in opening statements. *The Southern Communication Journal, 54*, 330-349.

Berger, C. R. & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research, 1*, 99-112.

Bickman, L. (1971). The effect of social status on the honesty of others. *Journal of Social Psychology, 85*, 87-92.

Bradac, J. J., Hemphill, M. R., & Tardy, C. H. (1981). Language style on trial: Effects of "powerful" and "powerless" speech upon judgments of victims and villains. *Western Journal of Speech Communication, 45*, 327-341.

Brown, B. L., Strong, W. J., & Rencher, A. C. (1973). Perceptions of personality from speech: Effects of manipulations of acoustic parameters. *Journal of the Acoustical Society of America, 54*, 29-35.

Brown, B. L., Strong, W. J., & Rencher, A. C. (1974). Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America, 55*, 313-318.

Burgoon, J. K., Birk, T., & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human Communication Research, 17*, 140-169.

Burgoon, J. K., Buller, D. B., & Woodall, W. G. (1996). *Nonverbal communication: The unspoken dialogue.* New York: Harper & Row.

Crassweller, P., Gordon, M. A. & Tedford, W. H. Jr. (1972). An experimental investigation of hitchhiking. *Journal of Psychology, 82*, 43-47.

Darley, J. M. & Cooper, J. (1972). The "Clean for Gene" phenomenon: The effect of students' appearance on political campaigning. *Journal of Applied Social Psychology, 2*, 24-33.

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

Erickson, B. Lind, A., Johnson, B. C., & O'Barr, W. M. (1978). Speech style and impression formation in a courtsetting: The effects of "powerful" and "powerless" speech. *Journal of Experimental Social Psychology, 14*, 266-279.

Giles, H., & Street, R. L. Jr., (1985). Communicator characteristics and behavior. In M. L. Knapp & G. R. Miller (Eds.), *Handbook of interpersonal communication* (pp. 205-261). Beverly Hills, CA: Sage.

Goldberg, S. (1982). *The first trial: Where do I sit? What do I say?* St. Paul, MN: West.

Haynes, R. (1984). How to be persuasive in a criminal trial? *American Bar Association Journal, 70*, 62-66.

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*, 136-144.

Hosman, L. A., & Wright, J. W. (1987). The effects of hedges and hesitations on impression formation in a simulated courtroom context. *Western Journal of Speech Communication, 51*, 177-188.

Krauss, R. M., Apple, W., Morency, N., Wenzel, C., & Winton, W. (1981). Verbal, vocal, and visible factors in judgments of another's affect. *Journal of Personality and Social Psychology, 40*, 312-320.

LaCrosse, M. B. (1975). Nonverbal behavior and perceived counselor attractiveness and persuasiveness. *Journal of Counseling Psychology, 22*, 563-566.

LeVan. E. A. (1984). Nonverbal communication in the courtroom: Attorney beware. *Law and Psychology Review, 8*, 83-104.

Lindsay, R. C. L., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses. *Law and Human Behavior, 13*, 333-338.

McCroskey, J. C., Jensen, T., & Valencia, C. (1973). *Measurement of the credibility of peers and spouses.* Paper presented at the International Communication Association convention, Montreal.

McCroskey, J. C. & Mehrley, R. S. (1969). The effects of disorganization and nonfluency on attitude change and source credibility. *Speech Monographs, 36,* 13-21.

McCroskey, J. C. & Young, T. (1981). Ethics and credibility: The construct and its measurement after three decades. *Central States Speech Journal,* 32, 24-34.

Mehrabian, A. & Williams, M. (1969). Nonverbal concomitants of perceived and intended persuasiveness. *Journal of Personality and Social Psychology, 13,* 37-58.

Miller, G. R. & Hewgill, M. A. (1984). The effect of variations in nonfluency on audience ratings of source credibility. *The Quarterly Journal of Speech, 50,* 36-44.

Miller, G. R., Fontes, N. E., Boster, F. J., & Sunnafrank, M. J. (1983). Methodological issues in legal communication research. *Communication Monographs, 50,* 33-46.

O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom.* New York: Academic Press.

Pearce, W. B. & Brommel, B. J. (1972). Vocalic communication in persuasion. *Quarterly Journal of Speech, 58,* 298-306.

Pearce, W. B. & Conklin, F. (1971). Nonverbal vocalic communication and perception of a speaker. *Speech Monographs, 38,* 235-241.

Rosenberg, S. W., Kahn, S. & Tran, T. (1991). Creating a political image: Shaping appearance and manipulating the vote. *Political Behavior, 13,* 345-367.

Scherer, K. R. (1979). Voice and speech correlates of perceived social influence in simulated juries. In H. Giles and R. N. St. Clair (Eds.), *Language and social psychology.* London: Blackwell.

Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 136-198). Cambridge: Cambridge University Press.

Schweitzer, D. (1970). The effect of presentation on source evaluation. *Quarterly Journal of Speech, 55,* 33-39.

Sereno, K. K. & Hawkins, G. J. (1967). The effects of variations in speakers' nonfluency upon audience ratings of attitude toward the speech topic and speakers' credibility. *Speech Monographs, 34,* 58-64.

Smith, M. J. (1982). *Persuasion and human action.* Belmont, CA: Wadsworth.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech, 18,* 145-152.

Sunnafrank, M. (1986). Predicted outcome value during initial interactions: A refomulation of uncertainly reduction theory. *Human Communication Research, 13,* 3-33.

Williams, G. R., Farmer, L. C., Lee, R. E., Cundie, B. P., Howell, R. I., & Rooker, C. K. (1975). Juror perceptions of trial testimony as function of the method of presentation: A comparison of live, color, video, black and white video, audio, and transcript presentation. *Brigham Young University Law Review,* 375-421.

Wrightsman, L. S. (1987). *Psychology and the legal system*. Monterey, CA: Brooks/Cole.

Zaidel, S. F. & Mehrabian, A. (1969). The ability to communicate and infer positive and negative attitudes facially and vocally. *Journal of Experimental Research in Personality, 3*, 233-241.

Appendix A

Nonverbal Coding Form

Lawyer # _____

Coder # _____

Fluency

Not fluent	1 2 3 4 5 6 7 8 9	Fluent

Vocalic Behavior

Little Pitch Variety	1 2 3 4 5 6 7 8 9	Great Pitch Variety
Little Tempo Variety	1 2 3 4 5 6 7 8 9	Great Tempo Variety
Little Volume Variety	1 2 3 4 5 6 7 8 9	Great Volume Variety

Kinesic Behavior

Unexpressive Face	1 2 3 4 5 6 7 8 9	Expressive Face
Few Illustrative Gestures	1 2 3 4 5 6 7 8 9	Many Illustrative Gestures
Restrained Gestures	1 2 3 4 5 6 7 8 9	Dynamic Gestures

Physical Appearance

Poorly Groomed	1 2 3 4 5 6 7 8 9	Well Groomed
Young	1 2 3 4 5 6 7 8 9	Old
Short	1 2 3 4 5 6 7 8 9	Tall
Slender	1 2 3 4 5 6 7 8 9	Heavy
Facial Hair	1 2 3 4 5 6 7 8 9	No Facial Hair
Unattractive	1 2 3 4 5 6 7 8 9	Attractive

Appendix B

Credibility Scale

Complete this form for each attorney you view.

Incompetent	1 2 3 4 5 6 7 8 9	Competent
Cruel	1 2 3 4 5 6 7 8 9	Kind
Telling Truth	1 2 3 4 5 6 7 8 9	Not Telling Truth
Aggressive	1 2 3 4 5 6 7 8 9	Meek
Tired	1 2 3 4 5 6 7 8 9	Energetic
Introvert	1 2 3 4 5 6 7 8 9	Extravert
Warm	1 2 3 4 5 6 7 8 9	Cold
Accurate	1 2 3 4 5 6 7 8 9	Inaccurate
Certain	1 2 3 4 5 6 7 8 9	Uncertain
Expert	1 2 3 4 5 6 7 8 9	Ignorant
Knowledgeable	1 2 3 4 5 6 7 8 9	Uninformed
Vague	1 2 3 4 5 6 7 8 9	Precise
Open	1 2 3 4 5 6 7 8 9	Defensive
Untrained	1 2 3 4 5 6 7 8 9	Trained
Unjust	1 2 3 4 5 6 7 8 9	Just
Admirable	1 2 3 4 5 6 7 8 9	Contemptible
Timid	1 2 3 4 5 6 7 8 9	Bold
Unfair	1 2 3 4 5 6 7 8 9	Fair
Insincere	1 2 3 4 5 6 7 8 9	Sincere

P300 Scalp Distribution as an Index of Deception: Control for Task Demand[1]

J. Peter Rosenfeld[2], Archana Rao, Matthew Soskins, and Antoinette Reinhart Miller

Abstract

Participants (n=24) experienced a baseline Block 1: they saw their phone numbers presented in a series with 6 other phone numbers. They were to say "yes" to their phone numbers, "no" to others. They were asked to repeat the first 3 digits of the phone numbers aloud. In Block 2, LIE and CONTROL groups (both n=12) were formed: participants saw a series of dates (e.g., "Mar 9"), 14% of which were their birth dates. The LIE participants were asked to lie on 50% of the trials, and to repeat all stimuli aloud. The CONTROLs were to perform honestly in Block 2, and were asked to repeat all stimuli aloud, but a random half of the stimuli backwards. The aim was to equalize task demand between groups. The results were that for both scaled and unscaled P300 amplitude, there were no differences or interactions as a function of group, or block in comparisons of responses to honest, forwards-repeated stimuli ($p>.6$). For pooled Block 1-Block 2 honest responses vs Block 2 dishonest responses in the LIE group, there was a main effect of response type on unscaled amplitude (lie responses<true responses, $p<.03$). Conversely, there was no main effect in the CONTROL group of the forwards/backwards manipulation ($p>.15$). In scaled amplitudes, there were no interactions of group or response type with site ($p>.2$) in honest, forwards responses. Comparing all scaled LIE honest with dishonest responses in the LIE group yielded a significant interaction of response type x site, $p<.02$. Post-hoc ANOVAs, using just Cz and Pz showed a significant interaction in the LIE but not CONTROL participants. There were no P300 latency differences between groups or conditions. In an extended replication, reaction time data did not differ between LIE and CONTROL groups. The results continue to support the notion that a P300 profile, specific for deception, may be identifiable.

Introduction

We have previously reported that in various situations, the scaled scalp distribution (profile) of P300 amplitude differs from deceptive to truth-telling conditions, (Rosenfeld, Reinhart, Bhatt, Ellwanger, Gora, Sekera, & Sweet, 1998; Rosenfeld & Ellwanger, (1999), Rosenfeld, Ellwanger, Nolan, Wu, Berman, & Sweet, 1999). Johnson (1988, 1993) has argued that when the ERP profile differs from one condition to another, this is good evidence that the two conditions involve differing neurogenerator groups.

In the present study, we tried to construct an honest control group having task demands comparable to those of the liar group. Specifically, we used an autobiographical oddball paradigm in which

---

[1]This article is reprinted with permission from *The Journal of Credibility Assessment and Witness Psychology* taken from Volume 3(1) pages 1 – 22. Please refer to http://truth.boisestate.edu for copies for reproduction under the original copyright that follows:

[2]Correspondence regarding this article should be addressed to Dr. J. Peter Rosenfeld, Department of Psychology and Institute for Neuroscience, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208. EMAIL: jp-rosenfeld@northwestern.edu

participants saw a Bernoulli (randomized) series of seven, repeatedly presented dates, 14.3 percent of which were their own birth dates. In the Lie group, participants were told to respond dishonestly on a random half of the trials (of both oddball and frequent type), and to then repeat the stimuli aloud. (Only the first three letters of the month were repeated.) In the Control group, participants were told to respond aloud honestly on all trials, but to then repeat a random half of the stimuli aloud backwards, (the rest, forwards). Both groups had comparable task demands in the terms noted above, but one group responded honestly and the other dishonestly. Differing P300 profiles would not be simply attributable to differences in task demands.

In this study, there is a second set of evidence examined which bears even more directly on putative specificity of Lie profiles: On the block of trials where the Lie participants respond dishonestly on half the trials, there is the opportunity to compare the P300 profiles associated with honest and dishonest response trials. Since task demands are the same during the entire block within the Lie group, obtained profile differences would provide support for the specificity hypothesis. We looked for but failed to find such an effect earlier (Rosenfeld et al., 1999) using a different (match-to-sample) paradigm.

We note that Johnson's (1988, 1993) interpretation of the meaning of differing scalp profiles emphasizes the possibility of differing neurogenerator sets. There is another interpretation of the differing scalp profiles in two experimental conditions: It may be that the two conditions evoke different sets of components which differentially overlap the P300 which both conditions evoke in common (Donchin, Spencer, & Dien, 1997). Either interpretation implies that the brain works in a specific way during deception, and the evidence would become the first to support a specific lie response, said to be a "dream" by Lykken (1981). Such a finding would also be a step in the direction of elucidating brain systems involved in lying.

Why might one expect differing scalp distributions in Lie and Control groups if task demand is matched? We hypothesize that a participant who is lying, even though he/she was directed to do so, has some level of self-awareness on all deceptive trials; that he/she is engaging in a behavior on which society and authority figures frown. At least some participants may thus find themselves somewhat embarrassed at being observed during lies. More important, all Lie participants (and no Control participants) know they are lying as they lie, and probably engage in further lie-specific cognitions following the decision to lie as well as following the act of lying. These cognitions would pertain to knowledge of the mismatch between the true-correct answer versus the answer they produce on a lie trial. We hypothesize that the Lie condition, but not the Control condition, will generate brain activity related (at least) to both the additional cognitions following such mismatch experiences, as well as to self awareness of deception, and that P300 profiles may reflect these differences between Lie and Control conditions.

Differences between Lie and Control groups might also be expected on the basis of the latter's additional task: backwards repetition of stimuli. A comparison restricted to profiles of Lie and Control groups during their respective specific tasks could thus be confounded by the two task effects simultaneously operating: 1) honest vs. dishonest responding and 2) backwards vs. forwards repetition. We therefore ran both groups through two blocks of trials, one (Block 1) in which all participants behaved alike in responding honestly and repeating stimuli forwards, and a second block (Block 2) in which the Lie participants lied on half the trials with forwards repetition, and the Control participants responded honestly on all trials but repeated half the stimuli in a backwards manner. Thus in each group, we could compare departures in Block 2 from the benchmark/baseline condition of Block 1.

Method

Participants: The 24 participants (12 per group, 13 female, six of which were in the Lie group) were recruited from the department introductory psychology pool and were fulfilling a course requirement. All had normal or corrected vision.

Procedure: Following signing of consent form, instruction, and electrode attachment, participants were seated in a recliner such that a video display screen was in front of their eyes. The visual stimuli were presented on this screen every 6.0 s, a relatively long interstimulus interval required for verbal responding so as to allow the artifact associated with vocalization to dissipate prior to the subsequent trial. The trial began with the onset of pre-stimulus EEG baseline recording for 104 ms. The stimulus then appeared on the screen and endured for the remainder of the ERP recording epoch = 1944 ms (total epoch = 2048 ms). Immediately after clearance of the stimulus from the screen, the message "Please Respond" was presented and lasted 2 s. The participant was required to respond during this time.

There were two blocks of trials used in this study. In the first block (Block 1), the visual stimuli were participants' phone numbers (p = .14) and other phone numbers (p = .86), each repeated as many times(about 40) as the subject's phone number. Both Control and Lie participants were told to respond aloud truthfully and ordinarily in this preliminary block. The timing and parametric settings in this benchmark/baseline block were the same as in the actual test block (2) to be next detailed. In this second block (Block 2), the stimuli were the first three letters of a month, followed by a number from 1 to 31, e.g., MAR 9. Thus, birth dates could be formed. The participant then said "yes" or "no" signifying birth date or other date, respectively, and then immediately repeated aloud the three-letter symbol of the month.

In the Control group, the participants were (in Block 2) instructed to respond honestly "yes" or "no" and to then repeat these month symbol letters aloud backwards on approximately half the trials of both types (birth date, non-birth date). They were also instructed to try giving a random, as opposed to patterned, series of forward and backward responses. We suggested to these participants that we were interested in how well people can generate random sequences of responses while doing a foreground task. We also alerted them that if the computer detected patterned responding, the experiment would be re-started.

In Block 2, the Lie group participants were instructed to simulate malingered cognitive deficit as in Rosenfeld et al. (1998), by making dishonest "errors" on both trial types about half the time in response to the "Please Respond" message. They were told to generate a random, unpatterned series of deceptive responses, since the computer controlling the experiment could discern patterns, and that they would not "beat the test" if patterned responding was discerned, and the experiment would be re-started. Immediately after their "yes" or "no" response, they were required to repeat the first three letters of the month (in the normal, forwards order). Both groups were told there would be 45 presentations of birth dates randomly interspersed among 276 presentations of other dates; i.e., six dates each repeated 45 times. This was done in order to help them score close to the 50% target rate of deceptive or backwards responses. Following the response window (2.0 s) was a second 2.0 s period of no events prior to the start of the next trial. (Verbatim instructions are available on request from the senior author.) Table 1 presents stimulus-response combinations for both groups and both blocks, with abbreviations.

EEG recording and analysis: EEG was recorded with Grass P511k preamplifiers with gain = 100,000, and filters set to pass signals between 0.1 and 30 Hz (3db points). Electrodes (Ag - AgC1) were attached to Fz,Cz, and Pz referenced to linked mastoids with the forehead grounded. Impedances were maintained below 5000 ohms. EOG was recorded from a bipolar pair of electrodes above and below the eye. EOG signals > 80 uV led to trial rejection and replacement. Amplified signals were led to

12-bit A/D converters (Keithley-Metrabyte) sampling at 125 Hz, and the digitized signals led to a computer for on-line sorting, averaging, and storage. The computer programs (by the senior author) also controlled stimulus presentation, and performed off-line filtering and analyses.

In the present study, P300 determination is based on a standard baseline-to-peak method: The computer searches within each participant's average ERP within stimulus, paradigm and response categories (see Table 1), within a window which extends from 400 to 1000 ms post-stimulus for the 104 ms segment average (13 data points) which is most positive-going. From this segment average, the average of the first, pre-stimulus, 104 ms of the recording epoch is then subtracted. The difference defines unscaled P300 amplitude.

The midpoint of the maximally positive segment defines P300 latency. This is a typical method of measuring P300 (Fabiani, Gratton, Karis, & Donchin, 1987).

Table 1: *Abbreviation Summary of stimulus-response combinations:*

(a.) LIE Group

Test Block 1 (all forward honest responses)

OD1[L]: oddball stimulus, honest response

FR1[L]: frequent stimulus, honest response

Test Block 2 (all forward responses)

OD2-TRU: oddball stimulus, honest response

OD2-LIE: oddball stimulus, dishonest response

FR2-TRU: frequent stimulus, honest response

FR2-LIE: frequent stimulus, dishonest response

(b.) CONTROL Group

Test Block 1 (all forward honest responses)

OD1[C] oddball stimulus, forward response

FR1[C] frequent stimulus, forward response

Test Block 2 (all honest responses)

OD2-FOW: oddball stimulus, forward response

OD2-BAC: oddball stimulus, backward response

FR2-FOW frequent stimulus, forward response

FR2-BAC frequent stimulus, backward response

The method just described is done only with Pz recordings. For the Cz and Fz sites, the temporal boundaries of the maximally positive segment at Pz are used to define the window over which P300 amplitude is calculated. This procedure is utilized to be certain that the same neural process is sampled across sites for purposes of profile construction. It is typically used by researchers who focus on scaled P300 amplitude profiles (e.g. Ruchkin, Johnson, Grafman, Canoune, & Ritter, 1992).

For group analyses, P300 latency and amplitude were based on unfiltered averages for each participant. For display, averages were digitally filtered to pass low frequencies; 3db point: 4.23 Hz. For task-by-site interactions, average P300 amplitudes within each participant were filtered and then scaled using the vector length method (McCarthy & Wood, 1985): Within each group and/or stimulus/response condition, the average Fz, Cz, and Pz values for the condition/group were squared, and the square root of the sum of the squared values was used as a denominator by which individual Fz, Cz, or Pz values within the condition/group were divided.

It is noted that analyses are performed here on both scaled and unscaled data. To look at main effects of group, stimulus type, block, response type, and scalp site on amplitude, it is appropriate to look at unscaled data (McCarthy & Wood, 1985). However, to answer questions involving interactions with site, (the major questions here) McCarthy & Wood (1985) explained the need for analysis on scaled data. What the scaling accomplishes is the removal of possible amplitude differences between conditions, which may confound amplitude distribution differences. The scaling procedure in the present study removes main effects of group, paradigm, response type, and stimulus type, and allows meaningful interpretation only of interactions involving site. Thus, as recommended by McCarthy & Wood (1985), we report analyses on both scaled and unscaled data, as appropriate. (Latency need not be scaled).

Extended Replication: The above procedures were repeated one year later, with one modification, on two new groups of Lie and Control subjects, (N=10, 11 respectively): Interspersed randomly among the oddball and frequent trials were 20 probe trials. On these trials, the word "Go" appeared on the computer screen and all participants were instructed to press a response button as soon as possible thereafter. This allowed us to obtain reaction time (RT) data and compare RTs between Control and Lie groups. Such information could then support our contention of equalization of task demands between groups; (RT is frequently used to assess task demand.) The probe trial stimuli appeared with the same timing as the other stimuli. Although electrodes were attached as in the original study and ERPs recorded, the ERP analysis presented is based on the original experiment. The modified replication was analyzed here only for RT data.

Results

Note: The key quantitative results on scaled data are in sections E and F below, and in Figure 6. Other results are reported immediately below in sections A, B, C, and D.

A. Behavioral (original study): The mean numbers of responses in each stimulus-response category (see Table 1 for abbreviations) are shown in Table 2. There are six rows in each group and the numbers in the first row in the Lie group should correspond to those in the first row in the Control group, the second row in the Lie group with the second row in the Control group, and so on. The appropriate correspondences are close except for the fifth (second to the last) row, involving frequent stimuli (Lie = 101.58 vs. Control = 87.75). For the first four rows involving the oddball responses in both groups in both blocks, and the frequents of Block 1, there were no significant differences.

Table 2: *Average numbers (± SEM) of responses in each possible stimulus-response category. Table 1 and text define category abbreviations*

Row	Category	Number
Lie Group		
1	OD1[L]	24.67 +/- .97
2	FR1[L]	146.58 +/- 5.80
3	OD2-TRU	17.25 +/- .85
4	OD2-LIE	15.00 +/- .90
5	FR2-LIE	101.58 +/- 4.46
6	FR2-TRU	86.50 +/- 4.24
Control Group		
1	OD1[C]	25.67 +/- .99
2	FR1[C]	143.67 +/- 7.16
3	OD2-FOW	15.17 +/- .91
4	OD2-BAC	14.83 +/- 1.28
5	FR2-BAC	87.75 +/- 6.54
6	FR2-FOW	87.80 +/- 5.26

There were significant effects regarding the last two rows containing frequent stimulus data, however these will not be detailed since all ERP analysis will focus only on oddball trials; P300s in many participants on frequent trials in both groups were dubious. The present behavioral data indicate comparability between groups for oddball stimulus-response combinations; (the differences found for frequents were small though significant).

B. RT data (modified replication): Average RTs to probe stimuli within each subject were averaged to yield separate group means, for each of the two blocks. For the first block in which all subjects performed in the same manner, the mean RT (+/- SD) for the Control group was 1.109S (+ .3984) and for the Lie group was 1.305S (+ .2098). On this difference, $t(19) = 1.425$, $p = .17$ (ns). In the critical second block, the differences were similar: Control = 1.02 S (+ .3925), Lie = 1.221S (+ .1927); $t(19) = 1.47$, $p = .16$ (ns). These negative data suggest that the two

tasks did not impose differential demands on the two groups of subjects.

C. ERP data: Qualitative observations in grand average ERPs: In the first block, there should be no ERP differences between groups in response to either oddball or frequent stimuli, since both Lie and Control groups are behaving exactly alike in this block (see Table 1 and methods). Differences between groups in amplitude and latency of P300 did not, in fact, reach significance (see below).

For quality control purposes, Figure 1 shows superimposed Lie and Control grand averages for OD2-TRU and OD2-FOW trials (all honest, forwards responses in block 2). It appears that the P300 is reduced in the Lie group relative to the Control group. Figure 2 shows superimposed Lie and Control grand averages for OD2-LIE (dishonest, forwards) and OD2-BAC (honest, backwards) trials, and again, the P300s appear larger in the Control group.

Figure 3 shows superimposed OD2-TRU (honest) and OD2-LIE (dishonest) responses within the Lie group. The former set appears to have more positive P300 responses, especially at Fz and Cz. (The differences

would be more obvious if we chose, in the figures, to superimpose pre-stimulus baselines, which our P300 calculation algorithm does do. We present data in figures as they really are, i.e., with random-noise related baseline shifts.) In Figure 4, comparable superimpositions are shown within the Control group: OD2-FOW (forwards) vs. OD2-BAC (backwards). In this comparison, P300 in the latter category appears slightly more positive (which, again, would be more evident with aligned baselines)

D. P300 amplitude data analysis: Unscaled data: We restrict reporting of results to oddball trials, since it was frequently impossible to locate a clear P300 peak in the frequent averages within participants.

Figure 5 shows the group average, computer-determined P300 amplitude values as functions of site, group, block (1 vs 2), and stimulus-response combination. It appears that within the Lie group, there is little difference in amplitude or slope, between OD1-[L] and OD2-TRU amplitudes (both associated with honest responses), but that lying (OD2-LIE) produces a depression of amplitudes. In the Control group, the OD1[C] and OD2-FOW response curves are also aligned, and indeed do not appear to differ from comparable Lie group honest response curves just described. This is as predicted. However, in the Control group, the OD2-BAC amplitudes appear enhanced by the backward condition manipulation..

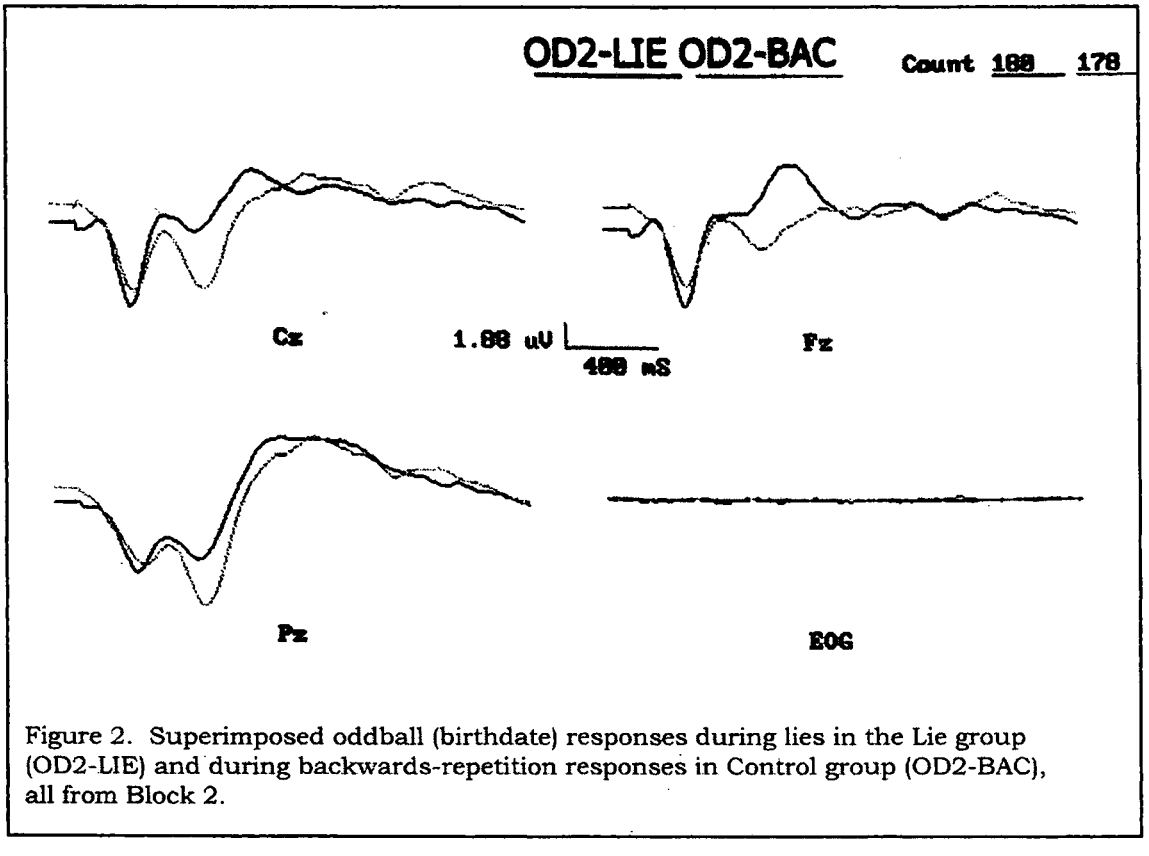


Figure 2. Superimposed oddball (birthdate) responses during lies in the Lie group (OD2-LIE) and during backwards-repetition responses in Control group (OD2-BAC), all from Block 2.

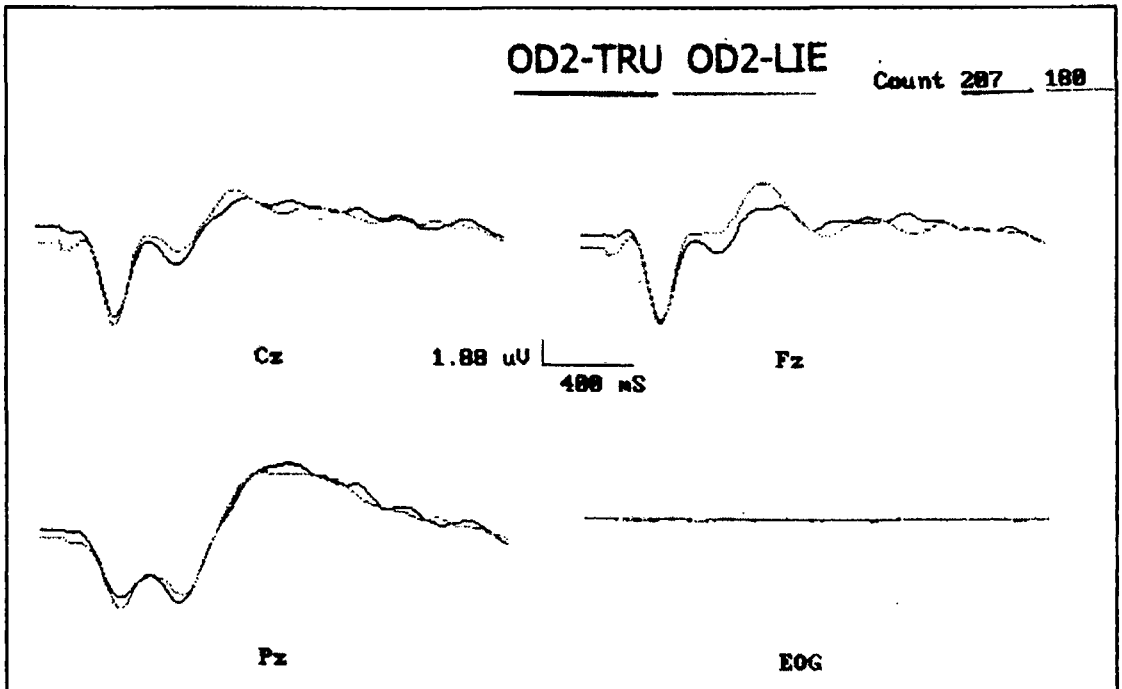


Figure3. Superimposed honest (thick: OD2-TRU) and dishonest (thin: OD2-LIE) responses, all from LIE group in Block 2.

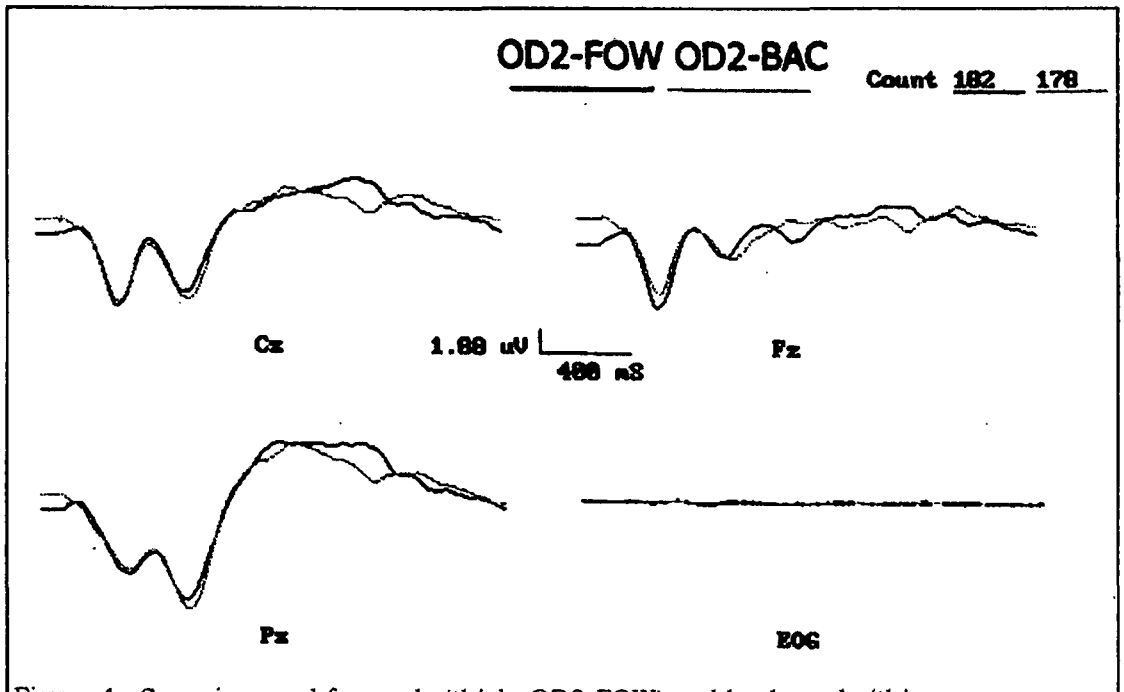


Figure 4. Superimposed forwards (thick: OD2-FOW) and backwards (thin: OD2-BAC) honest responses in Control group, Block 2.

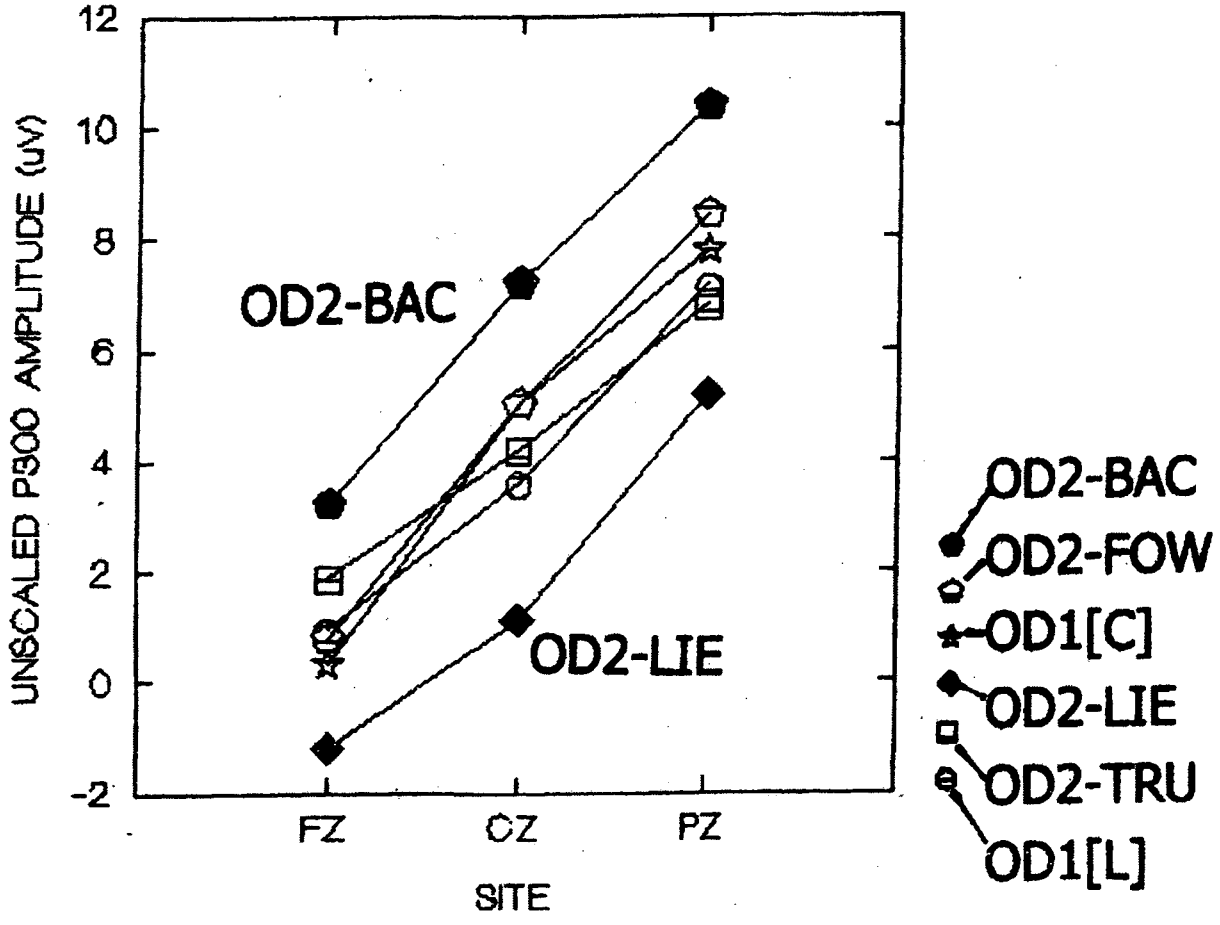


Figure 5. Averages of computer-determined, within-participant, unscaled P300 amplitudes (uV) as a function of site, paradigm, stimulus, and response type.

To obtain statistical confirmation of these effects, we first examined possible group and block differences during honest, forwards responses OD1[L], OD2-TRU, OD1[C], OD2-FOW. The sets of P300 amplitudes classified in this way were submitted to a 3-way ANOVA, with independent variables group (Lie vs. Control), site, and Block (1 vs. 2 for both groups).

The effect of group was not significant (p >.7). Neither was the effect of Block (p >.6). The effect of site yielded $F(2,44) = 134.34$, pg < .001 (pg is the Greenhouse-Geiser corrected probability in within-subject tests with df > 1. The correction is for sphericity effects. For df = 1 tests, the usual p-values will be reported.) The interactions were not significant, (p >.2), excepting the group-by-site interaction, which yielded $F(2,44) = 4.18$, pg < .04, reflecting the somewhat steeper slopes for honest, forwards Control curves than for the honest, forwards Lie curves in Figure 5. (As noted in the methods, without scaling or normalization of amplitudes, all interaction effects or lack of interactions, are possibly confounded and not simply interpretable).

To get at the effects of primary interest here, we compared each of the Block 2 special response types with their respective pooled truth-telling/forwards-repeating values. (Since the 3-way ANOVA described above

showed no differences between groups or block during truth-telling and forwards-repeating trials, the pooling was legitimate.) Thus we averaged OD1[L] and OD2-TRU to form OD-TRU, and we averaged the comparable Control data to form OD-FOW.

Within the Lie group, we then compared OD-TRU (honest) and OD2-LIE (dishonest) and examined site effects. The effect of site was $F(2,22) = 89.98$, $pg < .001$. The effect of honest vs. dishonest responses was $F(1,11) = 7.11$, $p < .03$, reflecting the lower value of averaged OD2-LIE responses in comparison with averaged OD-TRU (the pooled average of OD1[L] and OD2-TRU). The interaction of site and response type was not significant ($p > .4$). In the Control group, the effect of site was $F(2,22) = 73.36$, $pg < .001$. There was no significant effect of forwards versus backwards repetition ($p > .2$), despite the appearance of such a difference in Figure 5. Neither was the interaction of response type and site ($p > .6$) significant. Thus, although the dishonest response manipulation had a significant effect on unscaled P300 amplitudes in comparison with honest responses, the backwards repetition manipulation did not.

E. P300 Amplitude analysis; scaled data: group comparisons: In this section, we will comment only on interaction effects, since the scaling of data intentionally obviates main effects other than site effects, which are exaggerated (McCarthy & Wood, 1985). Figure 6 is the scaled equivalent of Figure 5, and shows scaled P300 amplitudes as a function of site, block, group, and response type. The figure suggests that all curves are similar except for the curve of the Lie group, during the second block, and only on dishonest response trials (OD2-LIE). We imply no interpretation of these scaled data which we simply here display (Figure 6) and describe (Ruchkin, Johnson, & Friedman, 1999).

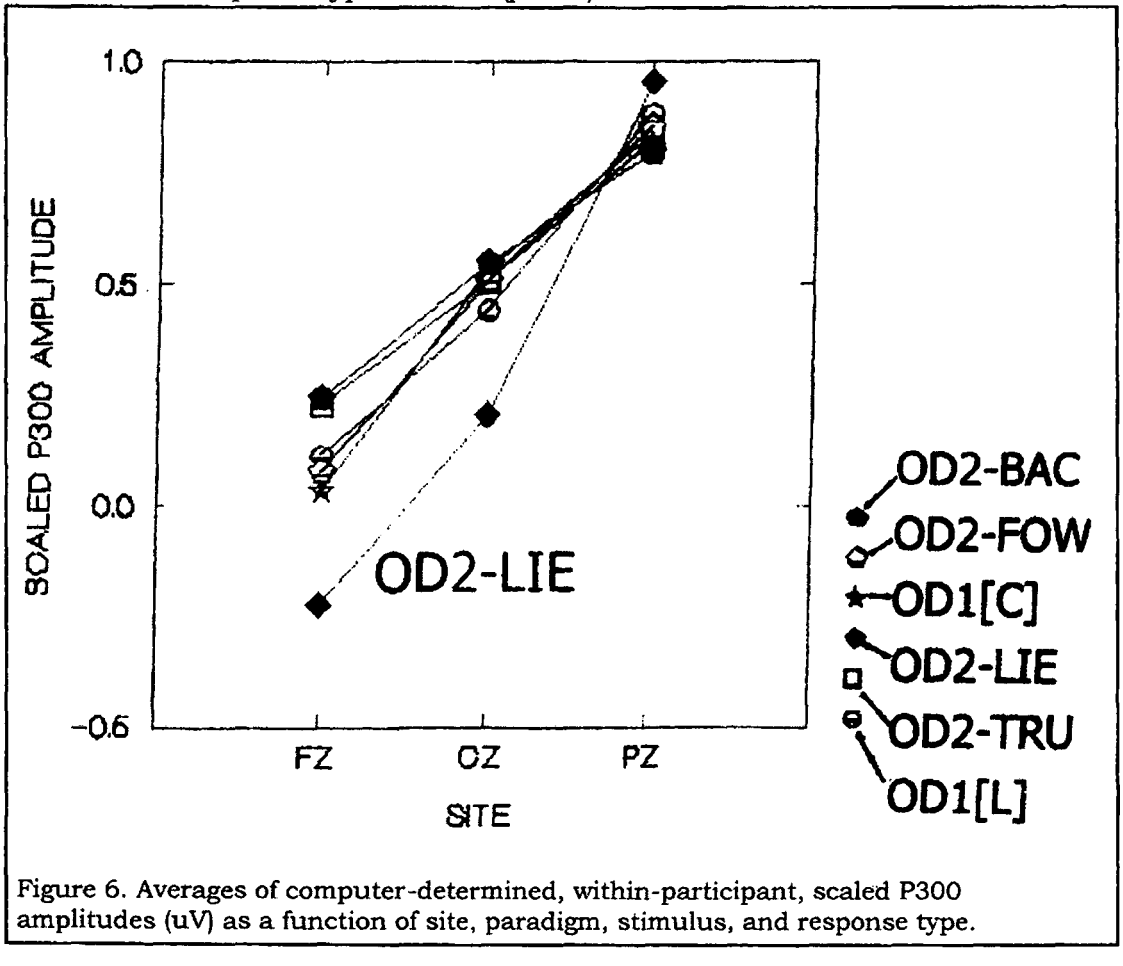


Figure 6. Averages of computer-determined, within-participant, scaled P300 amplitudes (uV) as a function of site, paradigm, stimulus, and response type.

Our statistical analysis approach with scaled data parallels the approach used with unscaled data. Thus the first analysis performed on scaled data was a 3-way ANOVA on all honest-responding, forward-repetition conditions, i.e., with independent variables: site, group, and block. The four response types separately submitted by group were OD1[L], OD2-TRU, OD1[C], and OD2-FOW. No interactions were expected, and none were found; (all $p > .2$).

Next, as with unscaled data, we combined the honest, forward response trials within each group to use as a benchmark-baseline with which to compare dishonest (Lie) or backwards (Control) responses. Thus, OD-TRU is the average of OD1[L] and OD2-TRU in the Lie group; OD-FOW is the comparable average within the Control group. Within the Lie group, a 2-way ANOVA on effect of response-type (OD-TRU vs. OD2-LIE) and site yielded a significant interaction of response type-by-site; $F(2,22) = 6.76$, $pg < .02$. Within the Control group, the comparable ANOVA on effect of OD-FOW vs. OD2-BAC with site also yielded a significant interaction; $F(2,22) = 10.6$, $pg < .001$. This was in contrast to what is suggested in Figure 6, where the scaled curves seem all alike (especially at Cz and Pz) except for the OD2-LIE (dishonest response) curve. It is noted (Figure 6), however, that whereas in the Lie group, the interaction shows (at Fz and Cz) a depression of OD2-LIE in comparison with OD2-TRU (honest vs. dishonest responses), in the control group, the OD2-BAC responses are (at Fz) slightly (though significantly) enhanced in comparison with the OD2-FOW curve. (These observations, again, imply no interpretation regarding relative activities or source strengths at the noted sites, but are meant simply to illustrate different kinds of interactions in Lie and Control groups; Ruchkin et al., 1999.)

We performed comparable ANOVAs, post-hoc, on data from just the Cz and Pz sites. In the Lie group, OD-TRU vs. OD2-LIE interacted with site, $F(1,11) = 24.32$, $p < .001$. However, in the Control group, OD-FOW vs. OD2-BAC did not interact with site ($p > .15$); neither did OD1-FOW vs. OD2-BAC ($p > .1$).

F. P300 Scaled Amplitude Analysis: Within Lie Group: The major comparison in this study is of the honest and dishonest response trials in Block 2 within the Lie group (OD2-TRU vs. OD2-LIE). This is because the task demands in the Lie group should be constant over trials within the block. A 2-way ANOVA on response type (honest/dishonest) and site did yield an interaction: $F(2,22) = 7.2$, $pg < .02$, as is evident also in Figure 6.

G. Latency Effects: Table 3 shows the Pz latencies of P300 for oddball responses in the two groups, segregated by response type. The Control group latencies are slightly greater than those of the Lie group (although the largest difference in row 1 of the table occurs prior to the group-generating manipulation). For both groups responding honestly and with forwards repetition in both paradigms, a 2-way ANOVA was performed on oddball latencies, with independent variables group and response type. There were no significant effects for group ($p > .2$), response type ($p > .5$) or interaction ($p > .4$).

Another 2-way ANOVA was performed on Pz latencies involving group and honest, dishonest, forwards, and backwards response types. Again there were no significant effects of group ($p > .4$), response type ($p > .6$), or interaction ($p > .6$). The present manipulations had no effects on P300 latencies, suggesting that stimulus processing task demands for the two groups did not differ, inasmuch as P300 latency has been associated with stimulus evaluation time (Fabiani et al., 1987; Johnson, 1988).

Discussion

We have shown previously (Rosenfeld et al, 1998; Rosenfeld et al., 1999) that the scaled scalp distributions (profiles) of P300 amplitude in deception conditions differ from those seen in simple truth-telling conditions. Since the scaled scalp amplitude distribution is independent of

Table 3: *P300 Pz Latencies ± SD*

Lie Group		Control Group	
Response Type	Latency (ms)	Response Type	Latency (ms)
OD1[L]	516 + 34.9	OD1[C]	550 +/- 53.9
OD2-TRU	518 + 80.2	OD2-FOW	528 +/- 47.3
OD2-LIE	518 + 49.6	OD2-BAC	539 +/- 44.8

Ivanov, & Mazzeri, 1988, Rosenfeld, Angell, Johnson, & Qian, 1991, Ellwanger, Rosenfeld, Sweet, & Bhat, 1996, Farwell & Donchin, 1991; Allen & Iacono, 1992.)

One could not say, however, on the basis of previous studies, that the profile seen in deceptive conditions represented neural activity specific to deception, itself, since, as reviewed in the introduction, deceptive and truth-telling conditions previously utilized also differed in task demand: the truth-teller had only to do his/her best on a simple task whereas the deceiver had to (additionally) keep track of his/her deception rate, and decide on each trial whether or not to lie.

The present study was designed to address these considerations in 2 ways: (1) allowing comparison of profiles between two groups (Lie and Control) in which we attempted to equalize task demand to the maximum possible extent, and (2) allowing comparison within the Lie group of profiles associated with honest versus dishonest response trials. Differing profiles in dishonest versus honest conditions would suggest different neurogenerator sets associated with each condition (Johnson, 1993; McCarthy & Wood, 1985). It may also be that the two conditions evoke different sets of components which differentially overlap the P300 which both conditions evoke in common (Donchin et al, 1997). In either case, however, the differing profiles indicate differing modes of brain function in each condition.

In fact, we found (Results, section F.) that scaled profiles differed in Lie group members during honest versus dishonest response trials. Since the task demand on the Lie group members was the same throughout the second paradigm task (i.e. during honest and dishonest trials), it is suggested that the significant interaction of response type (honest vs dishonest) by site provides evidence of differential modes of brain operation during the two kinds of trials, and that this effect is not confounded by task demand differences.

The Control group, like the Lie group also had to make a decision on each trial (whether or not to repeat a stimulus backwards), and had to track the same ratio of the two kinds of available responses (50-50). When scaled amplitude data from all three sites (Fz, Cz, Pz) were analyzed, this group also showed an interaction of site and response type (honest forwards repetition vs honest backwards repetition). However, the nature of the change from the forward repetition condition in the Control group was different than that seen in the Lie group. Indeed, if one considered only the Cz and Pz sites, then only the Lie group showed an interaction effect in the response type manipulation (response type x site) whereas the Control group showed no (response type x site) significant interaction. Similarly, in unscaled data from all three sites, significant main effects on amplitude were seen only in response to the honesty manipulation and not in response to the forwards vs backwards repetition manipulation. Thus the honesty-dishonesty manipulation had greater effects than the forwards-backwards manipulation (on

unscaled Fz, Cz, Pz amplitudes, and on scaled profiles at Cz and Pz) in this study.

Further evidence that group differences are not attributable to stimulus complexity aspects of task demand differences comes from the latency data: The P300 latencies did not differ between Lie and Control groups. Increases in task complexity involving greater stimulus processing demand from one condition to another are usually reported to increase P300 latency (and to decrease amplitude; Johnson, 1988).

It is also the case that in a modified replication of the present experiment in which probe stimuli were randomly inserted in place of date and number stimuli, there were no differences in reaction time to these probe stimuli between Lie and Control groups. This was further evidence of the comparability of task demand in these groups. We could not look at RTs to the other stimuli (as is often customary) because of the delayed response requirement necessitated by the need to avoid vocalization artifact. The probe stimuli, however, appeared in exactly the same time slots as did the other stimuli. They were more rare and when presented, were probably unexpected, as subjects most likely anticipated presentation of dates.

It is not surprising that in scaled profile data, the Control and Lie groups had differing profiles in Block 2 in comparison with their respective benchmarks. The two tasks are quite different in two ways, involving 1) honest (Control) versus dishonest (Lie) responses, and 2) trials with forward (Lie) versus backward repetition (Control). One could not say with certainty that by themselves, these differing profiles are due to honesty differences, repetition direction differences, or both. This is why we also used a first block with all participants responding honestly with forward repetition of stimuli. Since these profile data did not differ from the honest/forward repetition data in the second block, we pooled, within each group, the honest/forward response data from both blocks and used them as baseline/benchmarks with which to compare dishonest response profiles in the Lie group and backwards response profiles in the Control group. The manipulations within each group produced different scaled profile effects, in terms of shifts from the benchmarks as noted above, and we would attribute the effect in the Lie group to effects of deception.

This is consistent with the finding of different profiles for honest and dishonest responses within the second block of the Lie group, where within one block, different profiles were obtained. These effects might be attributable to deception specifically, since, as noted above, these Lie participants were all treated alike and the only difference between the cognitive states of Lie participants on trials involving honest vs. deceptive responses is this difference in response selection.

It is noted (Figure 6) that in the Lie group, the scaled OD2-LIE (dishonest response) curve is downshifted at Cz and Fz and upshifted at Pz relative to both the honest condition of Block 1 (OD1[L]) as well as to the honest response trials of Block 2 (OD2-TRU). It is also downshifted in comparison with all Control group curves at Fz and Cz, and upshifted at Pz. (We do not here intend to interpret the interactions on scaled data in terms of loci of cortical activity responsible for the interactions, as noted in the Results section, but only mean to describe unique features of the interaction in the Lie group.) These interactions strongly suggest that the lie response has a unique effect on brain operation. The fact that unscaled amplitudes are uniquely reduced in the Lie group during dishonest responses also supports a unique attribute related specifically with dishonest responses.

A question may be raised here regarding ecological validity. Our Lie subjects were not, in fact, lying in the way people do in the field. In our instructions to them, however, we repeatedly reminded them that when they would respond as if they were making errors, that in fact, they would know very well that these were not genuine errors, but lies. (One subject actually refused to complete the study at this point and was released.) Nevertheless, it remains a

limitation here that the subjects were executing directed rather than voluntary lies.

It was essential, in the design of this study, that there be no differences among the P300s associated with both blocks and groups during the honest responses. This requirement was mandated by our plan to pool honest, forwards responses so as to generate benchmark/baselines as described above. However, we also had application issues in mind: In any anticipated uses of these methods with real suspects in the field, it may be essential to have data from a control/baseline session, in which the suspect is known to be responding truthfully, with which to compare, in the same subject, data obtained during a test session in which the subject's (dis)honesty is to be ascertained. The present results in the Lie group which showed no differences between P300 distributions associated with truthful responses from both the first and second blocks, but differences between pooled truthful responses and dishonest responses, suggest that it should be possible to develop procedures, based on current group results, for future intraindividual diagnosis.

There is another implication regarding the data obtained from both groups during honest, forwards responding: One might have predicted differences between data sets obtained from the two blocks during honest, forwards responding on the basis of the fact that the first block utilized phone numbers as stimuli, whereas the second block utilized (birth) dates. A participant might have been expected to show different scaled amplitude profiles to these two kinds of stimuli on the basis of different cognitive processing of the two classes. Such differences were not observed. (Of course, such differences might be seen in data from other scalp sites.) This negative outcome suggests that the specific nature of the stimulus does not play a significant role in determination of profile shape in the present context: Rather, an autobiographal oddball stimulus yields a typical Pz > Cz > Fz profile which does not differ as a function of the specific nature of the stimulus, so long as an honest response occurs to the stimulus. Dishonest responses, however, affect the profile. We could have counterbalanced across participants the order of stimulus class used in the present study in order to control (unobtained) effects of differing stimulus classes. We chose not to counterbalance because while this counterbalanced design would have been easily implemented in the present laboratory analog, it would appear to present major problems in intraindividual field tests.

References

Allen, J., Iacono, W.G. and Danielson, K.D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology, 29*, 504-522.

Donchin, E., Kramer, A., & Wickens, C. 1986). Applications of brain event-related potentials to problems in engineering psychology. In M. Coles, S. Porges and E. Donchin (Eds.), *Psychophysiology: systems, processes and applications.* New York: Guilford.

Donchin, E., Spencer, K., & Dien (1997). The varieties of deviant experience: ERP manifestation of deviance processors in Van Boxtel, G.J.M. & Bocken, K.B.E. (Eds.), *Brain and Behavior: Past, Present, and Future,* Tilburg University Press, *p.* 116.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E (1987). The definition, identification, and reliability of measurement of the P3 component of the event-related potential. In *P.K.* Ackles, J.R. Jennings, & M.G.H. Coles (Eds.), *Advances in psychophysiology* Vol. 2, Greenwich: JAI Press.

Ellwanger, J., Rosenfeld, J.P., Sweet, J.J. & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology, 23,* 9-23.

Farwell, L.A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related potentials. *Psychophysiology, 28,* 531-547.

Johnson, R., Jr. (1988). The amplitude of the P300 component of the event-related potential. in P.K. Ackles, J.R. Jennings, & M.G.H. Coles (Eds.), *Advances in psychophysiology* Vol. 2 (pp. 69-138). Greenwich, Ct: JAI Press.

Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology, 30,* 90-97.

Kramer, A.F., Sirevaag, E.J., & Braune, R. (1987). A psychological assessmentof operator workload during simulated flight missions. *Human Factors, 29(2),* 145-160.

Lykken, D.T. (1981). *A tremor in the blood.* New York: McGraw-Hill.

McCarthy, G. & Wood, C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroenceph. Clin. Neurophysiol., 62,* 203-208.

Rosenfeld, J.P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology, 38,* 319-335.

Rosenfeld, J.P., Cantwell, G., Nasman, V.T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience, 24,* 157-161.

Rosenfeld, J.P., Reinhart, A.M., Bhatt, M., Ellwanger, J., Gora, K., Sekera, M., & Sweet, J. (1998). P300 Correlates of simulated amnesia on a matching-to-sample task: Topographic analyses of deception vs. truth-telling responses. *International Journal of Psychophysiology, 28,* 233-248.

Rosenfeld, J.P. & Ellwanger, J.W. (1999). Cognitive Psycholphysiology in Detection of Malingered cognitive deficit. *Forensic Neuropsychology: Fundamentals and Practice,* J.J. Sweet (Ed.), Lisse, Netherlands: Swets & Zeitlinger.

Rosenfeld, J.P., Ellwanger, J.W., Nolan, K., Wu, S., Bermann, & Sweet, J.J. (1999). P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model. *Int. J. Psychophysiol., 33(1),* 3-20.

Ruchkin, D.S., Johnson, R., Grafman, J., Canoune, H., & Ritter, W. (1992). Distinctions and similarities among working memory processes: an event-related potential study. *Cognitive Brain Research, 1,* 53-66.

Ruchkin, D.S., Johnson, R., & Friedman, D. (1999). Scaling is necessary when making comparisons between shapes of event-related potential topography. *Psychophysiology, 36,* 832-834.

Instructions to Authors

Scope

The journal *Polygraph* publishes articles about the psychophysiological detection of deception, and related areas. Authors are invited to submit manuscripts of original research, literature reviews, legal briefs, theoretical papers, instructional pieces, case histories, book reviews, short reports, and similar works. Special topics will be considered on an individual basis. A minimum standard for acceptance is that the paper be of general interest to practitioners, instructors and researchers of polygraphy. From time to time there will be a call for papers on specific topics.

Manuscript Submission

Manuscripts should be in English, and submitted, along with a cover letter, to Editor, American Polygraph Association, PO Box 10342, Ft. Jackson, South Carolina 29207 (USA). The cover letter should include a telephone number, return address, and e-mail address. Authors should also state clearly in the cover letter if they wish to submit their manuscript to a formal peer-review. The preferred method of manuscript submission is as an email attachment (MS Word, WordPerfect, or PDF format) with the cover letter included in the body of the email. Send to the Editor at:

Dsenter@sc.rr.com

Authors without Internet access may also submit manuscripts on computer disk along with 5 paper copies to the editorial address above. As a condition for publication, authors shall be required to sign a statement that all text, figures, or other content in the submitted manuscript is correctly cited, and that the work, all or in part, is not under consideration for publication elsewhere.

Manuscript Organization and Style

All manuscripts must be complete, balanced, and accurate. All authors should follow guidelines in the *Publication Manual of the American Psychological Association* (4th edition). The manual can be found in most public and university libraries, and can be ordered from: American Psychological Association Publications, 1200 17th Street, N.W., Washington, DC 20036, USA. Authors are responsible for assuring their work includes correct citations. Consistent with the ethical standards of the discipline, the American Polygraph Association considers quotation of another's work without proper citation a grievous offense. The standard for nomenclature shall be the *Terminology Reference for the Science of Psychophysiological Detection of Deception* included in this volume. Legal case citations should follow the *West* system.

Manuscript Review

A single Associate Editor will handle papers, and the author may, at the discretion of the Associate Editor, communicate directly with him or her. For all submissions, every effort will be made to provide the author a review within 12 weeks of receipt of manuscript. Articles submitted for publication are evaluated according to several criteria including significance of the contribution to the polygraph field, clarity, accuracy, and consistency.

Copy-editing

The Editor reserves the right to copy-edit manuscripts. All changes will be coordinated with the principal author.

Copyright

Authors submitting a paper to the American Polygraph Association (APA) do so with the understanding that the copyright for the paper will be assigned to the American Polygraph Association if the paper is accepted for publication. The APA, however, will not put any limitation on the personal freedom of the author(s) to use material contained in the paper in other works, and request for republication will be approved, if the senior author concurs.

Professional Copies

The senior author will receive ten (10) copies of the journal issue in which the article appears.