

Polygraph

VOLUME 34

2005

NUMBER 3

Contents

- | | |
|---|-----|
| Sexual History Disclosure using The Polygraph in a Sample
of British Sex Offenders in Treatment
Daniel Wilcox, Daniel Sosnowski, Brent Warberg, and
Anthony R. Beech | 171 |
| Polygraph Decision Rules for Evidentiary and Paired-Testing
(Marin Protocol) Applications
Donald J. Krapohl | 184 |
| Credibility Assessment Methods for the New Century
Donald J. Krapohl and John R. Timarco | 193 |
| Positions of Truthfully Answered Controls on Control Question
Tests with the Polygraph
Murray C. Cullen and M.T. Bradley | 197 |
| What does the Plethysmograph Indicate?
L.A. Geddes | 210 |

Sexual History Disclosure using the Polygraph in a Sample of British Sex Offenders in Treatment

Daniel Wilcox, Daniel Sosnowski, Brent Warberg, and Anthony R. Beech¹

Abstract

This pilot study on Sexual History Disclosure Examination was the first of its kind undertaken in the U.K., in that 14 adult male sex offenders were polygraphed after following completion of around 140 hours of Probation-based sex offender treatment. Of this sample eight were convicted of child molestation, four of indecent exposure and two for indecent assaults on women. Substantial increases in numbers of admitted victims and offenses were determined when comparing polygraph disclosure results with previously obtained data from all other available sources including information gathered in treatment. Participants also reported much earlier onset of offending behavior as well as a wider range of deviant paraphilic sexual interests than previously known and documented and perhaps most importantly a wider range of victims than previously known, i.e., 93% of the sample reported committing both contact and non-contact offenses, 50% of the sample reported committing intra-familial and extrafamilial offenses, and 29% of the sample admitted committing offenses against both children and adults after the polygraph examination. It is suggested that a larger study needs to be undertaken to investigate the usefulness of this approach with sexual offenders in the U.K.

Introduction

The use of the polygraph in Great Britain has only recently been given serious consideration as a potentially helpful technique for assessing, treating and monitoring sexual offenders (Salter, 1997, 1998; Wilcox, 1999). Middleton, Wilcox, and Sosnowski (1999) reported on the first British trial they completed using Specific Incident and Monitoring and Maintenance examinations as well as Sexual History Disclosure Testing (SHDT) which produced encouraging initial findings (Wilcox, Sosnowski, & Middleton, 1999a; 1999b; 2000). They identified important offense-related and offense-specific behaviors relating to the men polygraphed. For example, a man convicted of a sexual offense admitted a primary sexual attraction to a further victim in his family, although he had previously denied this over months of treatment. Another man convicted for a sexual offense had denied substantial

alcohol misuse, which was thought to be a significant feature in his offending, but admitted to its continued excessive use when polygraphed. Other men convicted of sexual offenses acknowledged violating probation conditions that were believed to be directly associated with risk of sexual recidivism (Wilcox et al, 2000).

Without doubt, British work in this area has been substantially influenced by the more extensive employment of polygraphy in the United States (Abrams & Abrams, 1993; Abrams, Hoyt, & Jewell, 1991; Abrams, & Ogard, 1986; Ahlmeyer, Heil, McKee, & English, 2000; Edson, 1991; English, Pullen, & Jones, 1996). Importantly, polygraph research findings have also shown significant crossover from known sexual offenses to other sexually deviant acts. The term 'crossover' addresses whether individuals had committed different categories of offenses from those on which official information was held

¹ Address for correspondence: Dr Daniel T Wilcox, Wilcox Psychological Associates, 9/10 Frederick Road, Edgbaston, Birmingham B15 1JD, Tel: 0121 454 8222, Fax: 0121 454 2999, email: daniel-wilcox@btconnect.com. Daniel Sosnowski, S.O.S. Services, GA., U.S.A. and University of Birmingham, U.K., Brent Warberg, S.O.S. Services, GA., U.S.A, Anthony R. Beech, University of Birmingham, U.K.

(Abrams & Abrams, 1993; Ahlmeyer et al, 2002; Edson, 1991; Franklin, 1995; Heil, Ahlmeyer, McCullar, & McKee, 2000; Heil, Ahlmeyer & Simons, 2003; O'Connell, 1998). Examples include: whether those known to have committed offenses against girls had also committed offenses against boys (or vice versa); whether those who had committed offenses inside the family had also committed offenses outside of the family (or vice versa); whether those who had committed offenses against children had also committed offenses against adults (or vice versa); whether those who had committed non-contact offenses had also committed contact offenses (or vice versa). By examining disclosures among large numbers of sexual offenders who have been systematically questioned about deviant interests, research (Abel & Rouleau, 1990) has been able to provide information about the crossover rates between even individual paraphilias.

Post-conviction sex offender polygraph research (English, Jones, & Patrick, 2002; English, Jones, Patrick, & Passini-Hill, 2003; English et al, 1996; Wilcox, 2000; Wilcox & Sosnowski, 2001; 2002) has particularly accelerated in the last decade. Polygraph studies have consistently evidenced higher numbers of sexual offenses and greater numbers of victims when convicted sexual offenders are tested.

The first use of the polygraph with sex offenders in the UK was undertaken in the fall of 1999 in the West Midlands Probation Service area in Central England (Middleton et al, 1999). The results gave indications that providers (therapists, probation officers) were able to obtain more detailed information about the offenses, further disclosures about past sexual offending and additional data about short-term risk and non-compliance with probation conditions. These findings led to the decision for a more extensive polygraph trial to be undertaken, specifically focusing on the Sexual History Disclosure Examination

(SHDE, see Appendix 1²) (Grubin, Parsons, Sosnowski, & Wilcox, 2002; Wilcox, 2002; Wilcox, Sosnowski, Middleton, & Grubin, 2002).

SHDEs are designed to obtain more detailed sexual histories about offenders. This procedure provides a better prospect for developing a more focused treatment approach (see the Association for the Treatment of Abusers guidelines, 1993). Abrams, S. (personal communication, December 8, 1998) also reported that a SHDE affords a good opportunity to demonstrate increases in the rates and ranges of sexual offense disclosures in comparison with other approaches (Wilcox, 2002). Therefore the aim of this research was to carry out the first trial of the polygraph, on a group of U.K. sex offenders who had already engaged in substantial specialized treatment, in order to investigate whether further offense related information could be obtained over and above that known from official data sources.

Method

Participants

The participants were all on Probation Orders and had been attending a sex offender treatment program for between three months and just over two years. These men had completed between 64 and 246 hours of group work (with a mean of approximately 141 hours). Twenty-five offenders had volunteered for SHDEs that had been administered in June 2000. One additional participant from the earlier October 1999 trials was included in the study as he was the only offender given a SHDE at that time. All of the men were living in the community, the majority independently. All of the participants who attended SHDE polygraph appointments in June 2000 (13 participants) and October 1999 (one participant) were included in the study. Therefore, the total number of participants reported in study was 14.

² Appendix 1 was developed further to the polygraph examiner's involvement as a member of the PCSOT Sub-Committee with the American Polygraph Association and has been employed in other polygraph research (Wilcox, Sosnowski & Middleton, 1999b). The form has been designed to address all of the paraphilic interests referenced in the DSM-IV (APA, 2000) as well as exploring other areas related to aberrant sexual behavior.

The average amount of time the participants had been in treatment was 13 months at the time of testing. The offenders ranged in age from 21 to 80 years and were all of white European ethnicity (mean age = 42). Based on the Risk Matrix 2000, (Thornton et al., 2003) the average risk level of this sample was in the medium to high range of risk. As for the type of offenses committed by men in the sample: four had convictions for intra-familial abuse (indecent assault); four had convictions for indecent assaults on children outside of the family, two of them with multiple similar offenses; four of the participants had convictions for indecent exposure³; the remaining two men had convictions for indecent assaults on adult women⁴.

Official record information

All comparative data known prior to the SHDE were collected on these participants through a thorough review of probation records, consisting of pre-sentence reports, group work and progress notes, psychometric tests and previous personal disclosure information reported. From these records, the following data were obtained: number of sexual offenses committed; age of onset of sexual offending; number and type of paraphilic (deviant sexual) interests disclosed; and crossover, in terms of the range of sexual offending behavior that was committed outside the parameters of the index offense, i.e., offenses being committed inside and outside of the family, offenses being committed against both males and females; offenses being committed against both children and adults.

Polygraph procedure

One of this report's authors (DS) is accredited by the APA to conduct Post Conviction Sex Offender Testing (PCSOT), additionally, the examiner is an approved trainer and a member of the APA, PCSOT

subcommittee. Therefore he conducted the SHDE.

The examinations took place in a quiet room where the West Midlands Probation Services Sex Offender Unit is based. The examinations took an average of two hours to complete. All participants were administered a single SHDE. Informed consent forms were completed and appropriate procedures were followed for all participants. Ordinarily, treatment workers and probation officers would have the opportunity to work closely with the examiner to develop appropriate polygraph questions concerning the SHDE. In this instance, owing to DS's limited availability, (having traveled from the USA for the specific purpose of conducting these trials), discussions with a particular treatment worker was brief given DS's time pressures.

Prior to each polygraph examination, DS had the opportunity to review the Probation files held on the individual at the Sex Offender Unit, focusing particularly on the pre-sentence reports for approximately a half-hour duration. In the pretest phase participants were shown how the polygraph examination apparatus worked and were given instructions concerning the test phase. This consisted of demonstrating how the offender's cardiovascular, respiratory and electrodermal skin resistance measures could be recorded and evaluated in relation to specific relevant questions agreed upon between the examiner and the polygraph participant. After this, the areas of inquiry raised in the SHDE (see Appendix 1 for SHDE questions) were fully explored with each participant and specific questions asked about the individual's past activities in these various areas of sexual involvement/ interest. After the SHDE each participant was given a polygraph examination to ascertain whether they were being truthful in answering the questions agreed upon during the pre-test.

³ All of the indecent exposer had multiple convictions with one demonstrating quite wide-ranging sexually deviant behavior including rape, indecent assault and attempted anal intercourse with a horse. Another indecent exposer had a history of extensive non-sexual criminal activities, mostly concerned with theft of property.

⁴One of these adult abusers was, at the time of this study, on probation for sending obscene mail to an elderly woman.

In the post-test phase, each participant was informed about whether they had passed the polygraph examination and given the opportunity to respond if a 'Deception Indicated' (DI) finding was made. In circumstances where deception or risk was indicated, participants knew that the results would be reported to their probation officers but as the polygraph trials were voluntary, no formal consequences could be applied for failing to attend or for responding deceptively to the polygraph questions. The only specific consequences that were uniformly agreed for all the men who volunteered to be polygraphed was that the information gathered would be shared with their field probation officers and their treatment providers. Participants were told that 'passing' the polygraph might lead to reduced levels of supervision or treatment if deemed appropriate by the professionals responsible. However, this did not occur for any participants as a result of polygraph testing. It should also be noted here that questions surrounding the test results and target selection were not the focus of the present enquiry.

Results

Ten of the 14 participants were found to be DI, two received No Deception Indicated (NDI) results, the other two individuals received No Opinion (NO) findings by the polygraph examiner. Those participants who failed the polygraph examination, e.g. received DI findings, were assessed by the examiner to be concealing information at the time of the polygraph examination, based on the review of the SHDE behaviors. The two individuals who passed the polygraph examination, receiving NDI results, were judged to have been appropriately disclosing about the SHDE items. The finding of NO for two further men occurred because one of them manifested a high level of distress that made the completion of the formal polygraph examination untenable and the other produced results that were best regarded as inconclusive as they could not safely be judged DI or NDI. An inconclusive result simply means that the polygraph examiner is unable to make a definite judgment. In this last eventuality, a second examination is usually conducted at a later date. Although the great majority of the participants were judged DI, a great many

disclosures were obtained from all of the participants.

The results of the polygraph examination were compared with official offense records on a number of offense-related variables as follows:

Reported number of offenses

A comparison was made between the number of contact and non-contact offenses that could be obtained from official probation records and a subsequent polygraph examination for each participant in the study. Wilcoxon non-parametric statistics were used due to the extreme range of scores reported by participants in the polygraph disclosure condition. Median and interquartile range (25% and 75%) scores, as well as mean and standard deviation (SD) values are reported in this section as these ordinal measures of central tendency and dispersion give a better idea both of a representative score compared to the means and SD values.

For contact sexual offenses (e.g., indecent assault, gross indecency), significantly more offenses ($z = 1.89$, $p < .05$, one-tailed test), were reported after the polygraph examination (mean = 81.9, SD = 188.4, median number of offenses = 4.0, interquartile range 3 to 83.5) compared to the official records (mean = 37.2, SD = 117.7, median number of offenses = 3.0, interquartile range 2 to 4.5). For non-contact offenses (e.g. voyeurism, exhibitionism), significantly more offenses ($z = 2.2$, $p < .05$) were also reported in the polygraph examinations (mean = 80.8, SD = 218.1, median = 5.0, interquartile range = 1.5 to 56) compared to official records (mean = 26.2, SD = 65.6, median = 3.0, interquartile range = 0 to 14.5).

Reported age at onset of offending

For this factor, parametric t-tests were used to compare age disclosed using the polygraph compared to age of onset of offending, officially known based on pretest information. Here we would note that this is the age that the offenders reported that they had started offending, rather than official figures, and therefore after the polygraph

examination gives a better picture of the real age that the sample were actually committing sexual offenses. Offenses were described with some contextual details included, although high frequency deviant behaviors including, for example, frottage and public exposure tended to be provided in terms of the individual's stereotypic pattern of offending with less situationally-relevant information available in some cases.

A significant difference ($t(12) = 2.9, p < .05$) was found between the participants mean known age at the time of onset of any kind of sexual offending (exhibitionism, voyeurism, indecent assault) based on official records and previous self report, where the mean age was 27.7, $SD = 18.4$, with earliest reported offending starting at 11 years of age. In comparison the first offense reported during

the polygraph examination was six years with a mean age for the sample of 13.4 ($SD = 3.8$).

Number and type of paraphilic interests disclosed

Paraphilic interests were examined in the interview because deviant sexual interests relate centrally to motivational (Finkelhor, 1986) and fantasy (Wolf, 1984) elements of the recognized treatment models employed in sex offender treatment work. Paraphilic interests were also examined because they indicate sexual preoccupation which is an important aspect of sexual interest (Thornton, 2002) and sexual self-regulation (Hanson & Harris, 2000). Table 1 indicates the types of deviant sexual interests either found from official sources or from the employment of the polygraph and whether the differences were significant using related t-tests.

Table 1: Paraphilic Interests Disclosed

Paraphilia	Number (%) from Probation records	Number (%) from the Polygraph examination	Sig. value (p)
Exhibitionism	3 (21%)	7 (50%)	ns.
Voyeurism	3 (21%)	9 (64%)	< .05
Obscene phone calls	0 (0%)	2 (14%)	ns.
Frottage	0 (0%)	7 (50%)	< .01
Fetishism	1 (7%)	6 (42%)	< .05
Sadism	1 (7%)	6 (42%)	< .05
Masochism	2 (14%)	4 (28%)	ns.
Transvestism/ Transsexualism	0 (0%)	4 (28%)	ns.
Public masturbation	5 (36%)	9 (64%)	ns.
Zoophilia	1 (7%)	3 (21%)	ns.

It can be seen from Table 1 that a significantly greater number of voyeuristic, fetishistic, sadistic and frottage paraphilic interests were reported after using the polygraph examination compared to those identified from probation records.

Crossover Rates

In terms of cross-over between contact offenses and non-contact offenses (e.g., exhibitionism, voyeurism, obscene telephone

calls, obscene mail) it was found that 93% of the sample (13 participants) reported both types of offenses whereas prior to polygraph examination only 29% of the sample reported both type of offenses (4 participants).

Based on Probation records, nine of the 14 participants (64%) had convictions for sexual offenses against children. Among them, three were known to have committed intrafamilial and extrafamilial offenses. Following polygraph examination, three more

of these nine convicted child abusers acknowledged intra- and extrafamilial abuse. This raised the crossover among known child abusers in the sample from 33% to 66%.

Probation records indicated that five of the 14 participants had committed rape/indecent assault offenses against adults. The SHDE's corroborated probation records and one further participant reported sexual offense(s) against adult victims following the polygraph examination. Among these six men, four reported previously unknown extrafamilial sexual offenses against female children following the polygraph examination⁵. Therefore, the adult/ child cross-over rate in this sample was 29%.

Information gathered on other potential disinhibiting factors

Information was also gathered from the polygraph examination regarding areas that are typically regarded as acute precursors to sexual offenses (Hanson & Harris, 2001; Hudson, Ward & McCormack, 1999; Pithers et al, 1988; Schwartz, 1995; West Midlands Probation, 2000). Here it should be noted that most participants endorsed items in multiple categories, specifically: 57% reported regularly using illicit substances including cannabis, as well as amphetamines and cocaine; 93% reported using, or abusing alcohol while on probation, 86% reported using explicit sexual material, i.e., magazines, videos, DVDs, or material downloaded from the Internet; 43% reported using prostitutes; 50% reported using nude bars; 43% reported committing adultery; 14% reported using telephone sex lines. It was also found that 57% of these men reported sexual stalking or "cruising" within the context of their pattern of offending. This type of information had not previously been known to the supervising officers of men in the sample.

Discussion

In this first British study of its kind,

the polygraph proved very beneficial as a means of obtaining more thorough, and in many cases hitherto unknown, details concerning the sexual histories of men attending a Sex Offender Treatment Program. A range of information was gleaned including offense-specific and offense-related data. Information was obtained concerning the range, type and number of sexual offenses the participants had committed, as well as their past level of engagement in activities that were accepted to be associated with risk of future offending. Significantly more information was disclosed following the polygraph examinations compared to that garnered from conventional sexual history information gathering methods. The types of extra information gathered included: a high level of substance/alcohol abuse; use of pornography; frequenting areas regularly used by children; trying to arrange private time with children; chronic or frequent masturbation to inappropriate fantasies; and improper use of the Internet. Therefore this study suggests that, through the employment of the polygraph a wealth of additional information regarding offense related behaviors can be obtained. This information would probably remain unknown to treatment providers and those supervising the offender in the community without the use of this instrument.

The findings of this study also replicate the results of earlier research identifying increases in offense reporting through employment of the polygraph (Ahlmeyer, et al 2000; O'Connell, 1998), in that the participants admitted to significantly more sexual offenses following the polygraph than was known from official records. Offenders reported over three times more contact offense victims and nearly five times more non-contact offenses than was previously known. Significant differences were also noted in of the numbers of paraphilias reported by polygraph participants, as compared with data obtained from all other sources available. Participants also acknowledged that their

⁵ Two of these men also had convictions for incest offenses while the other two adult abusers had no record of sexual offenses against children.

onset of offending occurred on average 14 years earlier than that officially identified for any type of sexual offense and 11.5 years previously for more serious offenses, excluding voyeurism and public masturbation.

Crossover rates were also an important area of investigation in this study. Historically, as noted by Abel and Rouleau (1990), it was believed that the offending patterns of sexual offenders were not particularly varied and as such, rather predictable. Abel and Rouleau challenged this notion and the findings of the current study illustrate the frequently diverse and robust nature of offenders' sexual behavior. Considerable evidence of crossover was determined in terms of committing contact and non-contact offenses, committing offenses against both genders and committing offenses against both adults and children. These crossovers were largely unknown in the sample until the polygraph was used, and therefore this technique has the potential to provide useful extra information in future risk assessment of sexual offenders.

It was also found that most participants endorsed multiple paraphilias (as indicated in Table 1). Therefore, we would suggest that these disclosures may represent high-risk behaviors for the participants in this research. Information was also gathered related to 'offense precursors' that are not directly associated with paraphilias. Currently the most useful framework that describes offense precursors is called a 'Decision Chain' (Ward, Loudon, Hudson & Marshall, 1995). A Decision Chain is a sequence of choices leading to an offense. Each choice is characterized in terms of the situation it took place in, the thoughts that made sense of and responded to the situation, and the emotions and actions that arose from these thoughts. Here it was found that most participants reported regularly using drugs, reported abusing alcohol and using explicit sexual material. A portion of the sample also reported using prostitutes, nude bars, committing adultery, or telephone sex lines. All of these additional findings can be seen as offering potentially important information for those involved in the community supervision of these offenders.

Limitations of the study

The limitations of the study include the small sample size, which hinders the generalizability of the findings. Also, we would note that all participants were volunteers, which again needs to be considered in determining the usefulness of these data. Another finding of this study was the absence of meaningful consequences for failure to participate; half of those who agreed to take part in the study did not. Their original decisions may have been motivated by a wish to demonstrate their innocence or a belief that they could deceive the polygraph examiner. However, as there was a period of several weeks between the time that the volunteers agreed to be polygraphed and the dates the examinations were actually administered, there is a question as to whether the non-participant rate may have been due to concerns about events that had transpired which would have exposed them to risk of detection. Of course, they may have simply opted not to participate. All of these concerns speak to potential biases within the sample, and thus the results should be interpreted with caution. Additionally, the method of sampling was self-selection insofar as the participants in the study were volunteers. Although this method does not yield a representative sample, it was the only realistic option available for conducting this preliminary research. Importantly, the participants were medium to high risk (Hanson & Thornton, 2000) sex offenders who had all been convicted of sexual crimes. As such, these findings should be considered as less applicable to individuals who fall within the low or very high risk ranges.

Implications for field practice

Sosnowski and Wilcox (2000) judged that unless meaningful sanctions are imposed, such as more intensive supervision, restrictions on activity within the community or financial liability for failure to participate, then the potential for making effective use of the polygraph might be seriously compromised. Clearly, many individuals have a great deal invested in concealing activities associated with their offending. As such, mandatory testing may be the most appropriate way to overcome this serious

problem. Importantly, Harrison and Kirkpatrick (2000) have noted that where sex offenders have been obliged to take the polygraph on a routine basis in a community setting, this factor positively influenced progress in treatment as well as offender behavior while on probation. Finally, Laws, Hudson, and Ward (2000) viewed the employment of the polygraph as providing sexual offenders with an improved potential to desist from offending in the future.

Future directions

Further research needs to be carried out in order to investigate whether the findings that there is a substantial increase in the numbers of admitted victims and offenses, an earlier onset of offending behavior, a wider range of deviant paraphilic sexual interests, and higher reported levels of crossover offenses (i.e., intra and extrafamilial offending, adult and child offending, and contact and not contact offending) when using the polygraph, compared to other sources of information. We would suggest that such a study should

investigate these findings in a larger, more representative sample of sexual offenders. We would also suggest that the additive benefits of the polygraph may also be more clearly shown if the SHDE is administered with and without the polygraph on matched samples of offenders in order to clearly demonstrate that it is the polygraph that is producing this extra information rather than the skill of the examiner in using the SHDE.

Acknowledgments

The authors would like to thank two anonymous referees and the editor for their helpful comments on an earlier draft of this paper.

Authors Notes

The substantive results of this study and additional findings of Wilcox (2002) are concurrently being published in Wilcox and Sosnowski (2005).

References

- Abel, G.G., & Rouleau, J.L. (1990). The nature and extent of sexual assault. In W.L., Marshall, D.R., Laws, & H.E., Barbaree (Eds.) *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 9-21). New York: Plenum Press.
- Abrams, S., & Abrams, J.B. (1993). *Polygraph testing of the pedophile*. Portland, OR: Ryan Gwinner Press.
- Abrams, S., Hoyt, D., & Jewell, C. (1991). The effectiveness of the disclosure test with sex abusers of children. *Polygraph*, 20, 204-213.
- Abrams, S., & Ogard, E. (1986). Polygraph surveillance of probationers. *Polygraph*, 15, 174-182.
- Ahlmeier, S., Heil, M., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 12, 123-138.
- Association for the Treatment of Abusers (1993). *Practitioner's handbook*. Lake Oswego, OR: Association for the Treatment of Abusers (ATSA).
- Edson, C. F. (1991). *Sex offender treatment*. Jackson County, OR: Department of Corrections.
- English, K., Jones, L., & Patrick, D. (2002). The polygraph plays a key role as a containment tool for convicted sex offenders in the community. *Polygraph*, 31, 240-253.

- English, K., Jones, L., Patrick, D., & Pasini-Hill, D. (2003). Sexual offender containment: Use of the postconviction polygraph. In R. Prentky, E. Janus, M. Seto, & A.W. Burgess (Eds), *Understanding and managing sexually coercive behavior. Annals of the New York Academy of Sciences*, 989, 411-426.
- English, K., Pullen, S., & Jones, L. (Eds.) (1996). *Managing adult sex offenders: A containment approach*. Lexington, KY: American Probation and Parole Association.
- Finkelhor, D. (1986). *A sourcebook on child sexual abuse*. Beverly Hills, CA: Sage Publications.
- Franklin, B. (1995). Polygraph disclosure testing. In D. W. Bragg (Ed.), *Polygraph handbook for probation/parole officers and treatment providers*, 23-51. Vienna, VA: Security Secrets, Inc.
- Grubin, D., Parsons, S., Sosnowski, D., & Wilcox, D. (2002, July). *The use of polygraphy in the management of sex offenders*. Presentation given at the 27th Annual Conference of the International Academy of Law and Mental Health, Amsterdam.
- Hanson, R.K., & Harris, A.J.R. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, 27, 6-35.
- Hanson, R.K., & Harris, A. (2001). The sex offender needs assessment rating (SONAR): A method for measuring change in risk levels. Available electronically from www.sgc.gc.ca/epub/corr/e200001a/e200001b/e200001b.htm. Please note this is an older version of SONAR and should not be used.
- Hanson, R.K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24(1), 119-136.
- Harrison, J.S., & Kirkpatrick, B. (2000). Polygraph testing and behavioral change with sex offenders in an outpatient setting: An exploratory study. *Polygraph*, 29, 20-25.
- Heil, P., Ahlmeyer, S., McCullar, B., & McKee, B. (2000). Integration of polygraph testing with sexual offenders in the Colorado Department of Corrections. *Polygraph*, 29, 26-30.
- Heil, P., Ahlmeyer, S., & Simons, D. (2003). Crossover sexual offenses. *Sexual Abuse: A Journal of Research and Treatment*, 15, 221-236.
- Hudson, S., Ward, T., & McCormack, J. (1999). Offense pathways in sexual offenders. *Journal of Interpersonal Violence*, 14, 779-798.
- Laws, P.R., Hudson, S.M., & Ward, T. (Eds.) (2000). *Remaking relapse prevention with sex offenders: A sourcebook*. Thousand Oaks, CA: Sage.
- Middleton, D., Wilcox, D., & Sosnowski, D. (1999). *Truth, lies and the polygraph*. Paper presented at the Annual Conference of the National Organization for the Treatment of Abusers (NOTA), York, England, September 1999.
- O'Connell, M.A. (1998). Using polygraph testing to assess deviant sexual history of sex offenders. *Dissertation Abstracts International*, 58, 3023.
- Pithers, W.D., Kashima, K.M., Cumming, G.F., Beal, L.S., & Buell, M.M. (1988). Relapse prevention of sexual aggression. *Annals of the New York Academy of Sciences*, 528, 244-260.

- Salter, A. (1997). *Sex offender assessment and risk management issues*. Paper presented at the National Organization for the Treatment of Abusers Annual Conference, Southampton, England, September 1997.
- Salter, A., (1998). *Sexual offenders: Assessment, risk management and treatment*. (workshop presentation). Specialized Training Services/ National Organisation for the Treatment of Abusers, London, June 1998.
- Schwartz, B.K. (1995). Characteristics and typologies of sex offenders. In B. Schwartz (Ed.). *The sex offender: Corrections, treatment and legal practice: Volume 2*. Kingston, NJ: Civic Research Institute, Inc.
- Sosnowski, D., & Wilcox, D.T. (2000). *Preliminary findings of Sexual History Disclosure Polygraph Examinations of British sex offenders* (poster presentation at the National Organization for the Treatment of Abusers Annual Conference. Dublin, Ireland, September 2000
- Thornton (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse. A Journal of Research and Treatment*, 14, 139-154.
- Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R, Friendship, C., & Erikson, M. (2003) Distinguishing and Combining Risks for Sexual and Violent Recidivism. In R. Prentky, E. Janus, M. Seto, and A.W. Burgess (Eds), Understanding and managing sexually coercive behavior. *Annals of the New York Academy of Sciences*, 989, 225-235.
- Ward, T., Loudon, K., Hudson, S., & Marshall, W.L. (1995). A descriptive model of the offense chain in child molesters. *Journal of Interpersonal Violence*, 10, 453-473.
- West Midlands Probation Service (2000) – *Sex Offender Groupwork Programme*. Accredited Programme/Home Office, UK.
- Wilcox, D.T. (1999). Polygraph testing. *NOTA News*, 30, 29-31.
- Wilcox, D.T. (2000). Application of the clinical polygraph examination to the assessment, treatment and monitoring of sex offenders. *The Journal of Sexual Aggression*, 5, 134-152.
- Wilcox, D.T. (2002) *Polygraph examination of British sexual offenders: A pilot study on sexual history disclosure testing*. Unpublished doctoral dissertation, University of Surrey.
- Wilcox, D.T., & Sosnowski, D. (2001, September). *A Review of UK Polygraph Research with Sex Offenders*. Presentation given at the Symposium on the Use of the Polygraph with Sex Offenders in Europe. Veldzicht Forensic Psychiatric Center, Balkbrug, Netherlands.
- Wilcox, D.T., & Sosnowski, D. (2002). Future European applications of the polygraph to sex offender assessment and treatment: A review of current UK research. *Probation in Europe*, 26, 6-7.
- Wilcox, D.T., & Sosnowski, D. (2005). Polygraph examination of British sexual offenders: A pilot study on sexual history disclosure testing. *Journal of Sexual Aggression*, 11(1), 3-25.
- Wilcox, D.T., Sosnowski, D., & Middleton, D. (1999a). Sex offenders, probation and the polygraph. *NOTA News*, 32, 18-23.
- Wilcox, D.T., Sosnowski, D., & Middleton, D. (1999b). The use of the polygraph in the community supervision of sex offenders. *Probation Journal*, 46, 234-240.

- Wilcox, D.T., Sosnowski, D., & Middleton, D. (2000). Polygraph and the community management of sex offenders. *Forensic Update*, 61, 20-25.
- Wilcox, D.T., Sosnowski, D., Middleton, D., & Grubin, D. (2002) British applications of the polygraph to sex offender work: An update. *NOTA News*, 41, 12-14.
- Wolf, S. (1984, November). *A Multi-Factor Model of Deviant Sexuality*. Paper presented at the Third International Conference on Victimology, Lisbon, Spain.

Appendix 1: List of Types of Behavior Discussed in SHDE

1. Sexual play with children (playing doctor or show me games)
2. Masturbation, How frequent?
Inappropriate Fantasies
Type of Fantasy
3. Masturbation in public (outdoors, restrooms, in vehicle)
4. Masturbation to pornographic material (XXX films, magazines)
5. Masturbation with under clothing/garments (stealing of clothes for masturbation)
6. Voyeurism (peeking/watching for sexual purposes)
7. Exhibitionism (exposing your sexual anatomy to others)
8. Incest (any sexual activity with a family member)
9. Homosexual behavior (any sexual activity with same sex, childhood or adult)
10. Obscene phone calls, including prank or nuisance calls.
11. Frottage (rubbing up against or touching others for sexual purposes)
12. Molestation (any sexual contact with minors as adult)
13. Aggravated Molestation (penetration of minor with penis, finger, or other objects)
14. Setting fires (for sexual purposes)
15. Prostitution (paying for sex with women)
15a. Prostitution with males
16. Stalking (following another person without their consent)
17. Phone sex calls (900, random, or someone known)
18. Sodomy (mouth to penis, vagina, or anus)
19. Transsexualism (thoughts or interest in wanting to be opposite sex)
20. Transvestitism (activities involving dressing in opposite sex clothes)
21. Sadism (deriving sexual pleasure from another's pain or humiliation)
22. Masochism (deriving sexual pleasure from receiving pain)
23. Taking photographs/videos of minors for sexual purposes.
24. Fetishism (sexual arousal from objects, --underwear, feet, shoes, vibrators
25. Placing objects into anus for sexual arousal.
26. Urolagnia (use of urine for sexual excitement--golden showers)
27. Coprophilia (use of feces for sexual excitement--brown showers)

28. Cruising (driving, walking searching for a sexual target)
29. Arousal to odors (any odor associated with sexual arousal)
30. Necrophilia (sexual contact with dead animals or people)
31. Animal sex or cruelty to animals.
32. Sexual victimization (if you have been sexually abused)
33. Grooming a minor or minor's guardian/parent with sexual intentions.
34. Contributed to the delinquency of minors--use of alcohol, drugs, shelter, porno)
35. Visiting areas where children frequent in order to have sexual contacts/fantasies.
36. Threesomes
37. Nude Bars
38. Sexual contact with fruit/vegetable
39. Other
40. Adultery
41. Hurt anyone during a sexual experience
42. Others not listed

Polygraph Decision Rules for Evidentiary and Paired-Testing (Marin Protocol) Applications

Donald J. Krapohl¹

Abstract

Most field research tends to find that the polygraph is better able to identify liars than truth-tellers, which has given rise to claims of an unfair bias in polygraphy. Prevailing polygraph decision rules came out of law enforcement, which places a high value on correctly identifying criminals for interrogation with an acknowledged cost of incorrectly interrogating some percentage of innocents. While useful in criminal investigations with high base rates and low consequence for false positive errors, these decision rules are not optimal for evidentiary applications, where false positive outcomes have much higher costs. A proposed set of Evidentiary Decision Rules and traditional Investigative Decision Rules were compared for correct decisions and the balance of errors using scores from four research projects. Evidentiary Decision Rules were found to produce more correct decisions, fewer inconclusives, and a more balanced accuracy than the traditional Investigative Decision Rules. Results suggest that Evidentiary Decision Rules may be better for courtroom and paired-testing (Marin Protocol) settings.

Polygraph decision rules constitute the critical final stage in the numerical evaluation process. These decision rules, in the form of cutting scores, establish the proportions of errors (false positive and negative) and inconclusive decisions for a given technique. It is recognized that decision rules of any imperfect diagnostic tool, including the polygraph, reflect the values of the user and that compromises are required to establish which cutting scores are used. In polygraphy, decision rules have almost universally been formulated by the developers or advocates of the various polygraph techniques, but there has been little discussion of instituting different decision rules for different applications. This is unfortunate, as certain decision rules that work well in one setting may have unacceptable error rates when used in another setting.

Consider the decision rules reported by Light (1999) for single-issue criminal testing. Decisions from blind scoring of charts using

Light's decision rules were shown to produce good detection of deception in field cases, but it was shown to have a lesser ability to detect truthful examinees with the same type of cases by Blackwell (1998). A general principle of decision-making is that one cannot reduce one type of decision error without increasing another type of decision error: reducing false positives increases false negatives. Therefore, decision makers must formulate rules that maximize the types of correct decisions that are most useful, limited only by their tolerance of an elevated rate for errors of the other type.

The decision rules reported by Light may work best when polygraph techniques are used to pare down the number of suspects, or to determine whether a suspect should be interrogated. The higher sensitivity to detect deception makes it unlikely that deceptive examinees would escape the additional focus that comes with a failed examination. The boost in accuracy for detecting liars comes at a cost: there is a commensurate loss in accuracy

¹This is one in a series of papers under the heading *Best Practices*. The opinions expressed in this article are those of the author, and do not necessarily represent those of the U.S. Government or the Department of Defense. Comments or reprint requests should be sent to: Donald Krapohl, DoD Polygraph Institute, 7540 Pickens Ave., Ft. Jackson, SC 29207, or by e-mail to dkrapohl@aol.com.

for detecting truthfulness. However, investigative polygraphy is not used as evidence to convict or incarcerate, only to help determine whether to keep the suspect in the radar. This creates a modest cost for false positives inasmuch as the individual was already a suspect when brought to the polygraph examination. If something like the polygraph never existed, the examinee would most likely have remained on the list of suspects anyway. Therefore, costs for polygraph false positives in this setting are small. If the cost of a false positive were higher in relation to that of a false negative, however, Light's decision rules would have to be modified to address the shifting costs and benefits.

There are two obvious settings where Light's investigative decision rules are not optimal: polygraph examinations tendered as courtroom evidence, and those conducted under the paired-testing or "Marin Protocol" (ASTM, 2005; Marin, 2000, 2001). In evidentiary polygraphy, the possible consequences of a false positive error to the examinee are no longer simply more questioning or police attention. It is important to contrast the goals of the investigative and evidentiary processes. The role of the investigative process is to gather evidence. A false negative decision in an investigative examination interferes with the ability of police to collect this information because it reduces attention on guilty suspects. Investigative decision rules better serve the interests of investigators when they minimize false negative results. By way of comparison, the role of the courts is to weigh the evidence regarding a suspect's involvement in a given crime. We would argue that the interests of justice in this context call for polygraph decision rules that are neither inclined toward false positives nor false negatives, though some might disagree.

Likewise, for situations suited for the Marin Protocol, highly unbalanced accuracies could reduce its intended usefulness. For example, taking the decision accuracies reported by Blackwell (1998) for the single-issue ZCT, one can calculate the likelihood of having one of two paired examinations result in error or inconclusive. Blackwell's study had shown very different hit rates for truthful

(44.8%) and deceptive cases (92.3%). The likelihood of having two correct decisions is product of the two independent probabilities, in this case $44.8\% \times 92.3\%$, or 41.4%. If there is a 41.4% likelihood of arriving at two correct decisions in a paired-testing protocol, then there is a 58.6% chance of having something other than two correct decisions ($100\% - 41.4\% = 58.6\%$). Let us now assume that the accuracies had been balanced, and equal to the average accuracy from the Blackwell findings. This is calculated as $(44.8\% + 92.3\%)/2$, or 68.6%. If there is a 68.6% chance of getting a correct decision without inconclusive or error, then the chance of two of these occurring independently of one another is $68.6\% \times 68.6\%$, or 47.1%. With a 47.1% chance of two correct decisions, the probability of at least one error or inconclusive is 52.9% ($100\% - 47.1\% = 52.9\%$). Recall that the unbalanced accuracies of Blackwell's original decisions resulted in a rate of 58.6% for at least one error or inconclusive result, compared to a 52.9% for the balanced accuracies. Therefore, balanced accuracies can result in fewer inutile polygraph decisions, in this hypothetical example by 5.7%. In the interest of maximizing the value of the Marin Protocol, highly unequal accuracies should be avoided. Moreover, there would be great benefit in finding a method that reduces inconclusives and errors. These factors prompted the present exploration of decision rules.

There are four confirmed factors that contribute to unbalanced accuracies. The first relates to how examinees are inclined to respond to relevant and probable-lie questions. Research with field cases has shown a marked tendency for examinees who are deceptive to the relevant issue to produce stronger responses to relevant questions than examinees truthful to the relevant issue produce to probable-lie questions (Franz, 1989; Krapohl, 1998; Raskin, Kircher, Honts, & Horowitz, 1988). The net effect is that the total scores from deceptive examinees are on average further below zero than the scores from truthful examinees are above zero. In other words, there is an asymmetry in the response patterns of deceptive and truthful examinees. Therefore, cutting scores that are symmetrical around 0 (i.e., ± 6) detect

deceptive examinees more easily than they do truthful examinees.

A second factor adding to this effect is the “spot score rule” (SSR; Light, 1999). The SSR dictates that a case, even a single-issue case, is called DI (Deception Indicated) when the sum of scores across three charts from any one of the relevant/comparison question spots is below -2. In single-issue testing the examination can also be called DI if the sum of all scores is below -5. For a decision of No Deception Indicated (NDI) in a single-issue exam each of the relevant questions must produce a total value that is positive, and the total of all questions must be at least +6. In other words, a DI call can be made from a single failed question, but an NDI call requires passing all questions plus achieving a +6 overall. The unequal standards for NDI and DI calls predictably make DI decisions much more likely than NDI decisions, hence the greater sensitivity to deception than to truthfulness.

The third factor that can contribute to the disparity in accuracy among truth-tellers and liars is the test question sequence. The order of the relevant and probable-lie questions has been found to produce a measurable effect on test scores, moving them in the negative direction when relevant questions immediately preceded by irrelevant questions rather than comparison questions (Cullen & Bradley, 2004; Krapohl & Dutton, *in press*). These findings point to problems with approaches such as the Reid Technique and the Army Modified General Question Technique (MGQT), where three of the relevant questions are immediately preceded by irrelevant questions.

The last factor involves test coverage. Some techniques use questions that cover different topics or different aspects of a crime within the same test. Called multiple-issue and multiple-facet techniques, respectively, these methods introduce the possibility that examinees may answer truthfully to some questions while lying to others. There are indications that these techniques are inherently inferior to single-issue techniques in terms of balance and overall accuracy. Senter (2003) conducted a careful analysis of data collected in various forms of the MGQT,

one of the multiple-facet formats. Senter concluded that “...the MGQT as an approach appears to be highly biased towards producing deceptive decisions. Thus, its use as a stand-alone diagnostic instrument may be problematic with respect to the production of false positive errors.” Further, field examinations where examinees were truthful to some questions and lying to others within the same test produces significantly poorer diagnostic scores than those where examinees had answered questions within a test either all truthfully or all deceptively (Raskin, Kircher, Honts & Horowitz, 1988; Podlesny & Truslow, 1993). Therefore, it would appear prudent to use only single-issue techniques for evidentiary or Marin Protocol examinations.

In the present investigation these four factors were considered, and they determined the types of cases that were selected and the type of decision rules that were applied to the scores. Only cases where the examinee had been entirely truthful or entirely deceptive had been included in the design. To avoid the negative shift in scores that Cullen and Bradley (2004) associated with question order, federal Zone Comparison Technique examinations were used. To compensate for the asymmetry in physiological responding to relevant and probable-lie questions, asymmetric cutting scores were combined with the two-stage decision rules of Senter (2001) with the aim of reducing inconclusives and balancing decision accuracy. It was expected that these steps would have two positive effects: They would produce decision accuracy that met the stringent requirements of the Marin Protocol (ASTM, 2005), and; the decision rules would correct the accuracy imbalance attendant to investigative polygraphy so to satisfy the expectation of fairness in evidentiary testing.

Method

Cases

Cases were randomly drawn from the archives of the DoD Polygraph Institute (DoDPI) except where noted. Because there was no exclusion of cases that had been used in other studies, it is possible that some overlap exists among samples taken from the DoDPI archives. All were single-issue field

cases² of criminal suspects where ground truth had been independently confirmed. They had been conducted with the Federal Zone Comparison Technique (Light, 1999). The examinees had been either truthful to every relevant question on the test or deceptive to every relevant question on the test. Two scorers evaluated 50 deceptive and 50 nondeceptive cases selected for the present study (Scorers 1 and 2). Another three scorers (Scorers 3 - 5) evaluated 61 deceptive and 39 nondeceptive cases as part of another project (Wygant, 2004). Scorers 6, 7 and 8 were polygraph examiners employed by a Florida law enforcement agency who evaluated 50 deceptive and 50 nondeceptive cases for an earlier study (Krapohl, Dutton, & Ryan, 2001). A research project by Capps and Ansley (1992) used 52 deceptive and 48 nondeceptive cases drawn from the archives of a DoD agency outside of DoDPI and scored by one examiner, labeled as Scorer 9.

Decision rules

Two types of decision rules were applied to the scores. The first was the traditional rules used for polygraph examinations by the U.S. government (Blackwell, 1998; Light, 1999), which will be referred to here as Investigative Decision Rules. According to these rules, a call of DI is made when the total of all scores is -6 or below, or if the sum of any one of the relevant questions is -3 or lower. Decisions of NDI require a positive total score for each relevant question and a grand total of +6 or greater for the sum of all spot scores. All other cases resulted in Inconclusive decisions.

The second method, referred to here as Evidentiary Decision Rules, begins with asymmetric cutting scores: if the grand sum of scores is -6 or lower, the call is DI; if the grand sum of all scores is +4 or greater, the call is NDI. In those cases where the grand sums

ranged from -5 to +3, the SSR is applied. For those cases, if a single relevant question has a sum of -3 or below, the decision is DI. All other cases are called Inconclusive.

The selection of these particular asymmetric cutting scores (-6, +4) was based on the data used to develop the Objective Scoring System version 2 (Krapohl, 2002). A reanalysis of those data showed that, on average, truthful examinees had only 78% of response differential to comparison questions as deceptive examinees had to relevant questions (R/C). For the present effort, if the standard -6 were accepted as the threshold for DI calls, an equivalent score for truthful cases would be just over +4 (-0.78 X -6). Therefore, the NDI threshold was rounded to +4.

Data reduction

Zeros were assigned where the examiner made notations that the tracing was uninterpretable due to artifacts. Only spot scores and total scores were analyzed. Decisions for each case were based solely on the scores and the two sets of decision rules. Examiner decisions were recorded, but not considered. Because of multiple comparisons, alpha for all statistics was set at .01.

Results

Investigative Decision Rules

Accuracy for Investigative Decision Rules is found in Table 1. The correct decision rate overall was 89.6% without inconclusives, which met the ASTM accuracy standard for the Marin Protocol of 86.0% or greater. However, the inconclusive rate was 22.1%, which exceeded the limit of 20%. Tests of proportion were applied to the polygraph decisions (Bruning & Kintz, 1997).

² By definition, single-issue polygraph examinations are those for which the relevant questions cover the exact same issue. For the cases used here, the third relevant question was often related to whether the examinee had knowledge about who committed the crime rather than whether the examinee had committed the crime. Adding this question increases the utility of the examination because it may lead to identification of the culprit by the police even when an innocent examinee is tested. It is known that there may be a loss of accuracy when an examinee is truthful to some questions and deceptive to others. Because the examinees in these cases were either deceptive to all relevant questions or truthful to all relevant questions, these cases met the veracity requirement of single-issue testing if not every technical aspect.

Of the 485 results possible on deceptive cases, 382, or 78.3% of them were correct, and this was greater than chance ($z = 9.13, p < .01$). A total of 259 correct decisions out of 415 opportunities were made on truthful cases for an accuracy of 61.5%, and this was also greater than chance expectancy ($z = 3.57, p < .01$). There was no difference in accuracy between truthful and deceptive cases ($z = 0.51, ns$) when inconclusives were removed, but proportion of correct decisions on deceptive cases was significantly higher than that of truthful cases ($z = 5.40, p < .01$) when inconclusives were considered.

The proportion of inconclusive results was greater for truth-tellers than for liars ($z = 4.98, p < .01$). Error rates were not significantly different between deceptive and truthful cases ($z = 1.55, ns$). Taken in sum, Investigative Decision Rules appear to make the examination sensitive to detecting lying, but there is a shortfall in its performance with truthful examinees. While error rates for truthful cases are not different from those of deceptive cases, truthful examinees do appear to suffer significantly more inconclusives.

Table 1. Decision accuracy using Investigative Decision Rules, in percentages.

Scorer	<u>Deceptive Cases</u>			<u>Truthful Cases</u>			<u>Overall</u>	
	Correct	Miss	Inc	Correct	Miss	Inc	Correct w/o Inc	Inc Rate
1	84.0	8.0	8.0	50.0	18.0	32.0	83.8	20.0
2	76.0	14.0	10.0	84.0	8.0	8.0	87.9	9.0
3	85.2	0.0	14.8	46.2	10.3	43.5	90.9	29.2
4	80.3	1.6	18.0	48.7	10.3	41.0	91.6	29.5
5	86.9	1.6	11.5	59.0	10.3	30.8	92.5	21.1
6	60.0	10.0	30.0	68.0	2.0	30.0	91.4	30.0
7	66.0	14.0	20.0	84.0	4.0	12.0	89.1	16.0
8	68.0	12.0	20.0	72.0	2.0	26.0	90.9	23.0
9	98.0	0.0	2.0	41.7	18.8	39.6	88.1	20.8
Average	78.3	6.8	14.9	61.5	9.3	29.2	89.6	22.1

Note. Inc = Inconclusive

Evidentiary Decision Rules

Table 2 shows the accuracy of the scorers when the Evidentiary Decision Rules are applied to the same scores. Of the 485 results possible on deceptive cases, 392, or 80.8% of them were correct, and this was greater than chance ($z = 10.09, p < .01$). A total of 333 correct decisions out of 415

opportunities were made on truthful cases for an accuracy of 80.3%, and this was also greater than chance expectancy ($z = 9.14, p < .01$). Correct decisions, as a percentage of all decisions, was 89.1%. The inconclusive rate was 9.6%. Both the accuracy and inconclusive rates met the Marin Protocol standard.

Table 2. Decision accuracy using Evidentiary Decision Rules, in percentages.

Scorer	<u>Deceptive Cases</u>			<u>Truthful Cases</u>			<u>Overall</u>	
	Correct	Miss	Inc	Correct	Miss	Inc	Correct w/o Inc	Inc Rate
1	82.0	10.0	8.0	76.0	15.0	9.0	86.3	8.5
2	74.0	20.0	6.0	90.0	6.0	4.0	86.3	5.0
3	85.2	0.0	14.8	66.7	10.3	23.1	93.3	18.9
4	78.7	9.8	11.5	87.2	2.6	10.3	93.0	10.9
5	86.9	3.3	9.8	89.7	5.1	5.1	95.5	7.5
6	70.0	16.0	14.0	78.0	14.0	8.0	83.1	11.0
7	76.0	16.0	8.0	86.0	6.0	8.0	88.0	8.0
8	76.0	18.0	6.0	76.0	12.0	12.0	83.5	9.0
9	98.0	0.0	2.0	72.9	12.5	14.6	92.9	7.5
Average	80.8	10.4	8.9	80.3	9.3	10.5	89.1	9.6

Note. Inc = Inconclusive

There were no statistically significant differences between the rates of correct decisions for truthful and deceptive cases when the Evidentiary Decision Rules were used ($z = 0.30$, *ns*). Inconclusive rates were not significantly different between deceptive and truthful cases ($z = 0.65$, *ns*), nor were error rates ($z = 0.13$, *ns*). These findings indicate that the Evidentiary Decision Rules did not share the bias of the Investigative Decision Rules in correct decisions or inconclusives.

Comparison of Decision Rules

Decision accuracy was compared between Investigative and Evidentiary Decision Rules. Evidentiary Decision Rules made significantly more correct decisions overall than did Investigative Decision Rules ($z = 4.69$, $p < .01$), and fewer inconclusives ($z = 6.23$, $p < .01$), but there were no differences in error rates ($z = 1.59$, *ns*). It appeared that most of the increase in correct decisions for Evidentiary Decision Rules over Investigative Decision Rules came from truthful cases ($z = 5.68$, $p < .01$) as there were no significant differences for deceptive cases ($z = 0.80$, *ns*) between the two sets of decision rules. In terms of errors, there were no significant differences for false negatives ($z = 2.00$, *ns*) or false positives ($z = 0.24$, *ns*) between the two types of decision

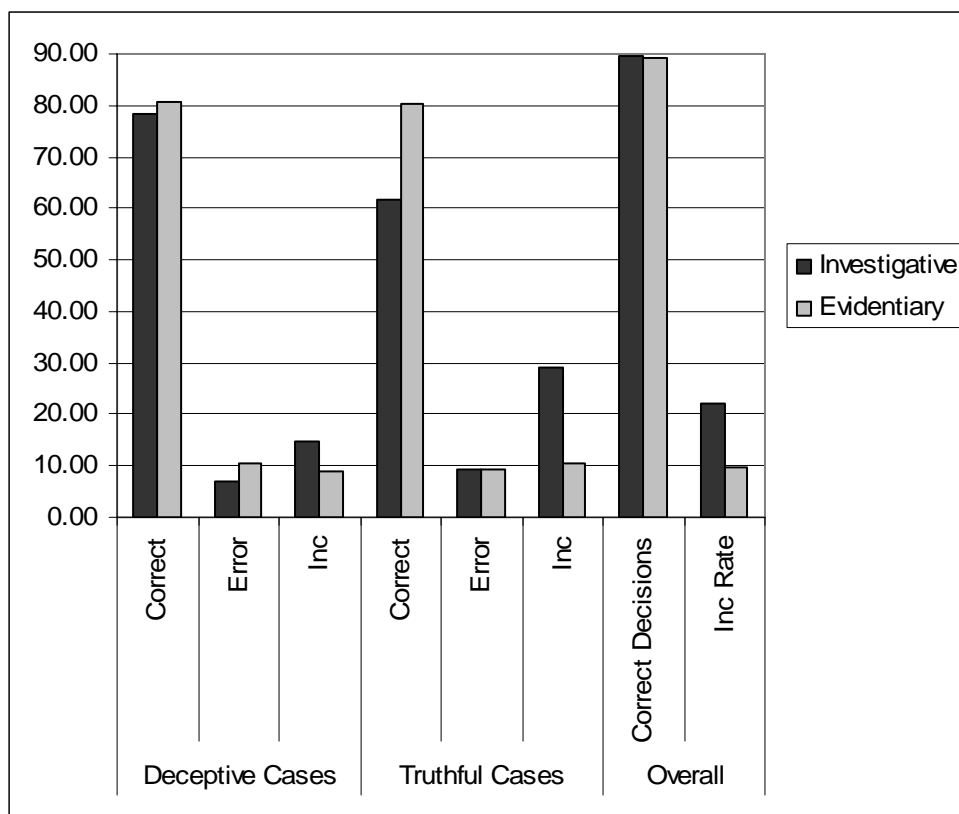
rules. Figure 1 compares accuracy between the decision rules.

Discussion

The Evidentiary Decision Rules performed well compared to the traditional Investigative Decision Rules. Evidentiary Decision Rules showed a marked improvement in detecting truthfulness, with no significant losses in detection of deception. It produced fewer inconclusives and more correct decisions with no significant differences in total errors. Evidentiary Decision Rules also had a better balance of correct decisions by virtue of improved accuracy for truthful cases that matched that of deceptive cases. Investigative Decision Rules made most of its correct judgments on deceptive cases, a finding that is consistent with those of most field studies, and these rules are perhaps best suited for law enforcement settings where they are most often applied.

Though the false negative error rates was not significantly different between the two sets of decision rules, it is worth noting that it did approach significance ($p = .046$). In terms of percentages, false negatives rose from 6.8% for Investigative Decision Rules to 10.4% for Evidentiary Decision Rules. While perhaps of practical importance, this difference falls just

Figure 1. Comparison of accuracy between Investigative and Evidentiary Decision Rules using the same four data sets. Correct Decisions Overall excludes Inconclusives.



short of the statistical threshold that would suggest that Evidentiary Decision Rules produced more false negatives than do Investigative Decision Rules. Future research should pay particular attention to these findings. If the difference attains statistical significance during replication it would suggest further refinements of the decision rules are necessary. A large-scale replication of this study is underway, and it will attempt to determine whether the lack of difference in false negatives is stable. In balance it is also important to note that along with the increase of 3.6% in false negatives found here, Evidentiary Decision Rules also produced an 18.8% increase in true positives and a 12% decrease in inconclusives.

The application of Investigative Decision Rules to the group data found that collectively the scorers did not meet the rigorous accuracy standards of the Marin Protocol. However, two individual scorers did

satisfy the Marin Protocol requirements: Scorers 2 and 7. When the Evidentiary Decision Rules were used, this number increased to seven of the nine scorers: Scorers 1, 2, 3, 4, 5, 7 and 9. The data suggest that the benefits of the Evidentiary Decision Rules may extend to more scorers than do the Investigative Decision Rules for meeting evidentiary or Marin Protocol standards. Scorers 2 and 7 met the Marin Protocol requirements under both sets of decision rules, a finding which tentatively suggests that the Evidentiary Decision Rules do not handicap those scorers who would have otherwise qualified for the Marin Protocol standard.

The Evidentiary Decision Rules are a natural extension of published findings from other workers in the field, however this was the first attempt to assemble them in the configuration reported here. The discovered trend would have been predicted from previous

research, but examiners should be mindful that this was a first look. Until replication, examiners should exercise caution in applying these rules. Relatively large data sets and nine independent examiners were used for this study, and there is reason to be optimistic that the findings will replicate.

A second limitation is that the proposed Evidentiary Decision Rules can only be applied to single-issue ZCT formats that use probable-lie comparison questions and require at least three charts. They are not appropriate for multiple-issue or multiple-facet examinations inasmuch as the Evidentiary Decision Rules rely initially on total scores. Total scores have little value when examinees can be truthful to some questions and deceptive to others within the same test. Caution is advised for using these rules with directed-lie techniques because their validity with these techniques has not been assessed. The findings can only be applied to the narrow conditions established for the type of data used in this study.

Summary

Traditional decision rules (+/-6 cutting scores and -3 spot scores) did not deliver the balanced accuracy and low inconclusive rate of the proposed Evidentiary Decision Rules (two-stage process: -6 and +4 cutting scores followed by -3 spot score for those case that would have been found inconclusive). Evidentiary Decision Rules for single-issue ZCT examinations appear to provide advantages for courtroom and paired-testing polygraph examinations, with no significant costs detected for that context. Most of the improvement was attributed to significantly better detection of truthfulness and reduction of inconclusives.

Acknowledgements

Thanks goes to Greg Anderson, Robert Drdak, Eugene Godfrey, Det. Larry V. Lingo, Glenda Leutwyler, Det. Herbert Metzgar, Kendall Shull, and James Wygant for scoring the polygraph cases, to Kay Williams and Betty Rodriguez for managing the data, and to John Schwartz, Stuart Senter Ph.D. and Dean Pollina Ph.D. for their helpful comments to an earlier draft.

References

- American Society for Testing and Materials (2005). E2324-04 Standard Guide for PDD Paired Testing. ASTM International.
- Blackwell, N.J. (1998.) PolyScore 3.3 and psychophysiological detection of deception examiner rates of accuracy when scoring examination from actual criminal investigations. *Polygraph*, 28(2) 149-175.
- Bruning, J.L., & Kintz, B.L. (1997). *Computational Handbook of Statistics* (4th Ed.). Addison Wesley Longman: Reading, MA.
- Capps, M.H., & Ansley, N. (1992). Comparison of two scoring scales. *Polygraph*, 21(1), 39-43.
- Cullen, M.C., & Bradley, M.T. (2004). Positions of truthfully answered controls on control question tests with the polygraph. *Canadian Journal of Behavioural Science*, 36(3), 167-176.
- Franz, M.L. (1989). *Technical report: Relative contributions of physiological recordings to detect deception*. Contract Number MDA 904-88-M-6612.
- Krapohl, D.J. (1998). Short report: Proposed method for scoring electrodermal responses. *Polygraph*, 27(1), 82-84.

- Krapohl, D.J., Dutton, D.W., & Ryan, A.H. (2001). The Rank Order Scoring System: Replication and extension with field data. *Polygraph*, 30(3), 172-181.
- Krapohl, D.J. (2002). Short report: Update for the Objective Scoring System. *Polygraph*, 31(4), 298-302.
- Light, G.D. (1999). Numerical evaluation of the Army Zone Comparison Test. *Polygraph*, 28(1), 37-45.
- Marin, J. (2000). He said/ She said: Polygraph evidence in court. *Polygraph*, 29(4), 299-304.
- Marin, J. (2001). The ASTM exclusionary standard and the APA "litigation certificate" program. *Polygraph*, 30(4), 288-293.
- Podlesny, J.A., & Truslow, C.M (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78(5), 788-797.
- Raskin, D.C., Kircher, J.C., Honts, C.R., & Horowitz, S.W. (1988). *A study of the validity of polygraph examinations in criminal investigations*. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040.
- Senter, S. M. (2001, July). *New approaches for increasing polygraph accuracy*. Presentation given at the annual meeting of the American Polygraph Association, Indianapolis, IN.
- Senter, S.M. (2003). Modified General Question Test decision rule exploration. *Polygraph*, 32(4), 251-263.
- Wygant, J.R. (2004). Analysis of scores to improve results. Unpublished manuscript.

Credibility Assessment Methods for the New Century¹

Donald J. Krapohl² and John R. Trimarco³

Introduction

The issue of lie detection was given prominence recently with the release of the report on the polygraph by the National Research Council (NRC, 2003). Prompted by Congressional interest in the government's use of the polygraph, the report was an exhaustive review of the available scientific evidence in the field to answer the question as to whether the polygraph was a valid tool. The NRC not only detailed the strengths and limitations of the polygraph, it also considered all other existing or potential approaches to the embryonic field of *credibility assessment*. Because their conclusions bear directly on technical approaches to credibility assessment that law enforcement can and does use, we recap here the relevant portions of the NRC report along with other recent research.

Before beginning with the technologies, it is instructive to first make a distinction between two principal approaches to credibility assessment. The first is what is called the recognition test. The basis of recognition testing is that examinees exposed to crime-relevant details will recognize them when they are imbedded among other similar-but-unrelated details. For example, suppose that a forger passed a check for \$921 at the Citizen's Bank on Saginaw Street last Wednesday. If the police had been careful to withhold the details of the crime from the press and the suspects, a potential suspect knowing the amount, place, and date of the check-passing is likely to have been involved in the crime. Using some of the credibility

assessment tools discussed later in this article, a culprit harboring this concealed information could be identified from the innocent suspects who do not know the crime details.

Deception testing is the second main approach to credibility assessment. Deception testing does not rely on whether the suspect recognizes incriminating details, but rather whether a direct question about the crime evokes a particular pattern of responses from the examinee. In the case of the polygraph, it is an identifiable response profile in the examinee's breathing, blood pressure, and skin conductivity. Other technologies may look for other signature patterns. The main advantages of deception testing over recognition testing is that it can be used when crime details have not been protected, when innocent persons may know them for legitimate reasons, or even when all of the crime details are not yet known to the police. Deception testing is used far more often than recognition testing because it can be applied in more situations. Conversely, recognition tests are generally easier to conduct than deception tests, and they are also more favored by many scientists. Some of the credibility assessment technologies are amenable to both recognition and deception testing, while others can only be applied to one or the other.

With the foundation provided above, let us now discuss the current and future approaches to credibility assessment.

¹ The opinions expressed in this article are those of the authors, and do not necessarily represent the views, opinions or policy of the U.S. Government or the Department of Defense.

² Donald J. Krapohl is employed at the Department of Defense Polygraph Institute where credibility assessment tools are investigated for the DoD. He can be reached at krpohl@jackson-dpi.army.mil or (803) 751-9107.

³ John R. (Jack) Trimarco is a retired FBI special agent, now President of a security firm, Jack Trimarco & Associates Polygraph / Investigations, Inc, in Beverly Hills, California. He can be reached at jtrimarco@aol.com or (310) 247-2637.

Current Technologies

Polygraph

The familiar polygraph is actually nothing more than a specialized physiological recorder. It doesn't detect lies per se, but rather registers bodily reactions associated with deception. Using validated protocols, a polygraph examiner can make reliable inferences of truth or deception based solely on the patterns of physiological responding. Guilty examinees, or those who are concealing their knowledge of details of the crime, produce telltale reactions that are different from those of truthful examinees.

The 2003 National Research Council report included an estimate of polygraph accuracy for criminal testing. Recognizing that real world conditions such as differences in criminal acts, examinee personality, and examiner proficiency, the NRC could not derive a single statistic to characterize polygraph accuracy. Instead they summarized polygraph accuracy as a range. For deception testing, it was from 81% to 91%, with a median of 86%. For recognition testing the range was 85% to 96%, with a median of 88%. The NRC concluded its assessment by reporting that "Some potential alternatives to the polygraph show promise, but none has yet been shown to outperform the polygraph." The release of NRC report has been followed by increased interest and federal funding in credibility assessment methods, with greater emphasis on emerging technologies.

Brain waves

For decades medical professionals have used brain waves to diagnose a variety of brain injuries, disorders and diseases. More recently scientists have used brain waves to examine cognitive processes such as sensation, perception, reading, and others, and it seemed logical that it might be applied to tests for deception or recognition. In the early 1990s a device was developed that focused on one particular type of brain wave called the P300 (a positive electrical potential occurring about 300 milliseconds after stimulus presentation). The P300 appears in the brain wave recordings of subjects when they were presented with a novel stimulus. It

occurred to researchers that crime details could be presented among other unrelated stimuli in a series of structured tests, and that a P300 would only appear among those subjects who recognized the crime details because they were "novel" in the context of the other stimuli. A small scale laboratory study was conducted that demonstrated that the P300 can be used in this manner (Farwell & Donchin, 1991) which showed an average accuracy was 88%. Since that time there have been attempts to use this technology in criminal investigations. Because the technology is restricted to recognition tests, and conventional polygraphy can also conduct recognition testing and is more widely available, brain wave testing in criminal matters has been extremely limited. However, federally funded research is currently underway using an exciting new paradigm to investigate the potential of brain waves in a deception test.

Voice Spectrum Analysis

Since first introduced in the early 1970s, voice spectrum analysis (VSA) has gained popularity among police as a low-cost and easy-to-use alternative to the polygraph. Fueled by an aggressive marketing campaign and manufacturer reports of success stories, VSA devices have grown to become second only to the polygraph in the number of field users. Research conducted over the last 60 years supports the contention that certain components in the voice, such as fundamental frequency, often change during stress. It is reasonable to conclude that computers and software could facilitate the analysis of voice samples provided by criminal suspects for evidence of stress during questioning, and thereby create a deception or recognition test.

This optimism appears to have been premature, however. While there are several commercial VSA devices on the market, the National Research Council concluded that there was no scientific evidence that any of them are valid. Moreover, the NRC was pessimistic that a voice-based device would ever be sufficiently valid as a lie detector. The NRC findings were not made in isolation, as numerous scientific studies of the commercial VSA devices have found they perform poorly as credibility assessment tools. Paradoxically,

even in the shadow of these conclusions new VSA devices are regularly being introduced to the market. Newer models purportedly not only detect lies, but have the ability to divine the subtle differences between white lies, defensive lies, and offensive lies, a capacity that outpaces even the most scientifically advanced expertise and technology if true. Other VSA claims include the ability to discern uncertainty, anticipation, even love. None of these assertions have independent support, and some fly in the face of established science. The failure of existing VSA technologies to demonstrate validity bodes ill for this approach to credibility assessment, but does not preclude the possibility of another method from emerging with the necessary scientific foundation.

Eye Movement

The movement of the eyes from one focal point to another is called saccades. Saccadic eye movements can now be tracked precisely using special equipment and

computer technology so that it is possible to determine where a person is looking and for how long. University scientists have discovered that the pattern of saccades is different between looking at pictures of familiar faces versus pictures of unfamiliar faces with an accuracy of 90% (Cohen, McConkie, Webb, Althoff, Holden & Noll, 1992). An effect has also been found between viewing familiar and unfamiliar scenes and objects, though with a lower accuracy. These discoveries are the basis of a new technology that can be used in recognition tests, though not for deception testing. Potential applications might include showing suspects a series of crime photos, only some of which relate to the crime under investigation, and automated “line-ups” for witnesses and victims to view. Work remains to be done before this technology is generally available, but it holds some potential for law enforcement applications in the coming years.

Table 1 summarizes the accuracies of the current technologies.

Table 1. *Accuracy of Existing Credibility Assessment Technologies*

	<u>Deception Tests</u>	<u>Recognition Tests</u>
Polygraph	86%	88%
Brain Wave	*	88%
Voice Spectrum Analysis	Near chance	Near chance
Saccadic Eye Movement	*	90%

* Technology is not yet designed for deception testing

Future Technologies

Thermal Image Analysis

The advent of better cameras and faster computers has given birth to a promising new approach to credibility assessment, thermal image analysis (TIA). TIA exploits the minute changes in facial skin

surface temperature to reveal physiological changes during deception. Several government and university laboratories have joined the race to develop a TIA method that can be used in a variety of applications; from rapid airport screening to detainee questioning, from criminal testing to applicant vetting. TIA does not require sensors that come in contact with the examinee, and it may

find applications where other technologies may be unsuited.

Because TIA is a new approach to credibility assessment, many issues require investigation: vulnerability to environmental factors or countermeasures, and the effects of health, race, age, metabolism, movements, cosmetics and others, to name a few. Nevertheless, thermal is one of the “hot” topics in the field of credibility assessment research.

Brain Imaging

Functional Magnetic Resonance Imaging (fMRI) is best known as a non-intrusive method for peering inside the body. It can also identify areas in the brain that are more active than the surrounding tissue. This potential may be useful in credibility assessment research. Because deception and recognition are cognitive processes with neural underpinnings, it is hoped that the fMRI will uncover specific regions of the brain that are uniquely involved in these processes. The basic research has started, and early findings have shown merit for recognition tests. As stated earlier in this article, recognition testing can already be done with several other technologies, and so the search has begun for an fMRI method of deception testing. There is great optimism that this line of research will eventually bear fruit, but there are problems with field applications that won't be easily overcome. Even if the fMRI is found to be extremely accurate in deception testing, the

significant purchase and operating costs of the multi-ton electromagnet along with a supporting staff would severely limit its field deployability. Nevertheless, fMRI may still serve an important role, such as pointing the way to other approaches to credibility assessment not previously considered, developing a cognitive model of deception, or even using it in high-stakes circumstances such as treason or death penalty cases. There is general agreement, however, that this approach to credibility assessment is a long way from practical use.

Conclusion

The tragedy of September 11, 2001 and the continuing Global War on Terrorism have given impetus to the search for a new “polygraph,” one that is more versatile and accurate, with greater speed and convenience. In a host of laboratories across the US scientists are searching for, devising, testing, disproving and proving an array of credibility assessment tools for the mission of protecting our public, troops, industry, transportation systems, and freedoms. This will be a large and long-term undertaking, and along the way we will see a few setbacks, false starts and snake oil gadgets. But the essential scientific breakthroughs will most certainly come as the best minds take on the problem. Today's investment in research will one day deliver better tools for those who are charged with the protection of our nation and our communities.

References

- Cohen, N.J., McConkie, G.W., Webb, J.M., Althoff, R.R., Holden, J.A., & Noll, E.L. (1992). *Detecting guilty knowledge using eye monitoring data*. Final Report to the U.S. Government.
- Farwell, L.A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy (“lie detection”) with event-related brain potentials. *Psychophysiology*, 28, 531-547.
- National Research Council (2003). *The Polygraph and Lie Detection*. National Academies Press: Washington, DC.

Positions of Truthfully Answered Controls on Control Question Tests with the Polygraph¹

Murray C. Cullen and M. T. Bradley²

Abstract

Students ($N = 120$) guilty of a mock crime, innocent and informed, or innocent and uninformed of crime details were examined by polygraph with an altered form of Control Question Test (CQT). Ambiguous, lie-engendering control questions were altered to form clear direct questions answered truthfully. When these control questions were positioned before crime-relevant questions, most guilty and innocent participants were correctly classified. Most participants were classed as guilty when crime-relevant questions were positioned before control questions. Lying to crime-relevant questions in the second position resulted in skin resistance, $F(2, 108) = 8.2$, and blood volume, $F(2, 108) = 6.1$, responses larger than Orienting Responses to initial control questions. Accurate detection depends on the position of control questions.

The present study manipulated the position of modified "control" questions in a Control Question Test (CQT; Reid & Inbau, 1977). It follows on the work of Bradley, MacLaren, and Black (1996) who introduced modified "control" questions to the CQT. The core of a typical CQT involves pairing "crime-relevant" questions concerning a specific crime with "control" questions. Physiological responses to adjacent pairs of "control" and "crime-relevant" questions are measured and compared. The question from each pairing that produces the larger response leads to the judgment of guilt if it is to the "crime-relevant" question, or innocence if to the "control" question.

"Crime-relevant" questions are direct and straightforward (Abrams, 1977). If a car was stolen, suspects may be asked, "Did you steal Mr. Doe's car?" Innocent suspects should show relatively small physiological responses in truthfully denying the theft. A guilty suspect through lying with a "no" should show a large physiological response.

"Control" questions in the usual form of a CQT are not as straightforward as "crime-relevant" questions. They are wide ranging, ambiguous, and in all probability elicit a deceptive response if a suspect answers "no" (Abrams, 1977). To illustrate, a question such as "Between the ages of 20 and 30 years of age did you ever steal anything from someone who trusted you?" puts some suspects in an uncomfortable position. They can confess to one or more incidents of this type and appear like a thief or they can deny such a general question but feel at least somewhat deceptive in their answer. In theory, this could be disturbing for innocent suspects and evoke strong physiological responses. They are in the process of being interrogated for a crime they did not commit and in that process they may be incriminating themselves over other issues. Guilty suspects are less disturbed because control issues are tangential to the crime under investigation. Their priority is the immediate one of evading detection on the questions relevant to the crime.

¹ This article originally appeared in the *Journal of Behavioural Science* in 2004, volume 36 (3), pages 167-176. Copyright 2005. Canadian Psychological Association. Reprinted with permission.

² Murray C. Cullen and M. T. Bradley, University of New Brunswick. Correspondence concerning this article should be addressed to M. T. Bradley, Department of Psychology, University of New Brunswick, P.O. Box 5050, Saint John, New Brunswick, Canada E2L 4L5 (E-mail: Bradley@UNBSJ.ca).

The CQT is controversial. Kircher and Raskin (1988) present data showing that the test is highly effective. Lykken (1981) claims the test makes errors, particularly false positive errors in which innocent suspects are found guilty.

Analyses by Lykken (1981) and Ben-Shakhar and Furedy (1990) argue that the CQT is fundamentally flawed. "Control" questions are not matched controls for "crime-relevant" questions in the scientific sense of the word and may not even be "comparable" in the normal language sense of a synonym. Part of this problem is that the test is transparent. It is obvious to both guilty and innocent suspects that "crime-relevant" questions are the ones to be passed to avoid the consequences of a trial and subsequent punishment. How, following one argument, could "control" questions be evocative of relatively strong physiological responses when they are clearly on issues tangential to the investigation?

Reviews (Iacono & Patrick, 1988; Saxe, Dougherty, & Cross, 1985) support Lykken's (1981) analysis. The reviews, however, are selective and greater familiarity with the literature (Honts, 1993) allows the contention that the test does work as designed. The controversy remains heated but unresolved. There has been no movement in the applied field to discontinue the use of the CQT despite the criticisms of some psychologists who think the test does not work.

Conflicting validity studies and differing theoretical conceptions have created an impasse. Bradley et al. (1996) examined empirically the elements of the CQT that are the subject of contention. They questioned the alleged necessity of "control" question characteristics. They created a different type of "control" question by eliminating ambiguity through making "control" questions direct and specific. The questions were phrased such that no innocent participant would lie. The following example illustrates how they phrased questions for a mock crime theft of \$20. A "crime-relevant" question was "Did you steal \$20?" A typical "control" question might be, "Have you ever stolen something of value from a friend?" The altered "control" question was, "Did you steal \$15?" In comparing the two

forms, the typical "control" question is broad, ambiguous, and the suspect may even be uncertain as to whether he or she is lying or not, depending upon the interpretation of the meanings of "value" and "friend." The modified "control" question, in contrast, is straightforward and unambiguous, especially in the context of the particular theft, and does not require a lie from an innocent participant. Further, the issue of question transparency was addressed by informing one group of innocent participants about the details of the crime.

The altered question form shares similarities with the Guilty Knowledge Test, or "GKT" (Lykken, 1959). The altered "control" question is directly comparable to an individual foil in a GKT when the innocent participant is unaware of the key item. When the participant is aware of the key item, the "control" question is like a foil in a Guilty Actions Test (GAT; Bradley & Rettinger, 1992).

Bradley et al. (1996) tested three groups of participants. A "guilty" group committed a mock crime and was required to be deceptive to "crime-relevant" questions. An "innocent yet informed" group knew all of the crime-relevant details but were not deceptive because they did not do the crime. An "innocent and uninformed" group was not deceptive and had no basis for distinguishing between "control" and "crime-relevant" questions. Guilty participants scored as deceptive. Participants in both innocent groups scored as truthful even though members of one group knew the details.

Bradley et al. (1996) presented their test with each "control question" in a pair always prior to the question about the crime. Consequently, order and content were confounded. It is not clear whether the results were due to question order effects or question characteristics. It is unlikely, given the nature of the modified "control" questions, that physiological responding was due to content, but in the logic of experimental design it is necessary to manipulate question order to examine this issue.

Order effects have been mentioned (e.g., Bradley et al., 1996) but not tested in the context of the CQT. Lykken (1959) in his

development of the GKT assumed that ORs to the first items in sets of questions would be stronger than subsequent responses to any other items regardless of their relevance to a crime. Therefore, the GKT was designed never to involve guilty information in the first position and never to include the first item in scoring the test. With the CQT, however, not all users have ensured that "control" questions always are prior to "crime-relevant" questions (e.g., Elaad & Elaad, 1994).

Ben-Shakhar and Lieblich (1982) manipulated the serial position of relevant items in a GKT paradigm, but only for positions following the first item. They found hit rates were higher when relevant items were placed at earlier rather than later serial positions in their lists. This finding was not replicated by Ben-Shakhar, Asher, Poznansky-Levy, Asherwitz, and Lieblich (1989). They found similar detection rates for early as well as late presentations of the relevant item within a series.

Bradley et al. (1996) assumed an order effect and placed "control" questions in the first position of a scored pair. Their intention was to take advantage of the OR to ensure that innocent participants would have larger responses to "control" questions than to subsequent "crime-relevant" questions even though truthful on all questions. ORs habituated over subsequent presentations but responses of innocent participants to initial questions still remained larger because of their position. The key experimental and practical issue resolved in the study was that physiological responses generated from guilty subjects deceptive to "crime-relevant" questions in the second position could exceed ORs to "control" questions in the first position.

In the present study, guilt condition and "control" question order were manipulated. The hypotheses were generated under the assumption that detection scores of participants depend jointly upon their actual guilt condition and the serial position of the crime-relevant items. With "crime-relevant" questions first, it was predicted that most participants, due to OR effects, would score towards guilt. If ORs are not preemptive of physiological reflections of other cognitive and emotional processes, guilt and knowledge were

predicted to have some potential to augment ORs. Therefore, with "crime-relevant" questions first, guilty participants and knowledgeable participants would be classed as guilty more so than innocent uninformed participants. With "crime-relevant" questions second, guilty participants were predicted to be accurately classed as guilty. That is, their physiological responses associated with lying would exceed the initial "control" question OR. Innocent participants even with "crime-relevant" knowledge were predicted to be accurately classed as innocent because they were not deceptive on the subsequent "crime-relevant" questions, and knowledge without deception would not be generative of a response to exceed initial ORs.

Method

Participants

One hundred and twenty Introductory Psychology student volunteers, 60 male and 60 female, took part in the study. Participants received a bonus point to add to their Psychology grade.

Apparatus

A Lafayette 750-566 field polygraph was used to record skin resistance responses (SRR), blood volume (BV), thoracic respiration responses (THR), and abdominal responses (AB). SRRs were measured by zinc-zinc chloride electrodes attached to the medial phalanges of the first and third fingers of the participant's right hand. Respiration was measured by two pneumatic tubes positioned around the thoracic area and the abdomen. Cardiovascular activity (a combination of heart rate and blood volume) was measured with a photoplethysmograph attached to the participant's second finger on the right hand during the interrogation.

Interrogators

Two graduate students alternated between serving as the lab assistant and the interrogator. Both of the graduate students had been trained by the second author of the present study while conducting prior studies under his direction.

Procedure

A consent form was given to each student interested in participating in the study. It contained information describing the study, as well as information on the potential risks and benefits of participation. An important feature of the consent form was the clear indication that participants could withdraw from the study at any point without penalty.

Individual participants were asked to go to the experimental area at an agreed upon time and report to a laboratory assistant. That assistant randomly assigned a file folder and gave the instructions in that folder to the participant. The instructions were for one of three conditions (guilty, innocent, or innocent and informed). The folder, which was retained by the assistant, also contained the polygraph examination questions in their predetermined order to be given later to the interrogator.

Each participant in the "guilty" condition read and carried out a set of instructions requiring him or her to go to a specific professor's office and: a) enter without knocking; b) remove \$20 out of a wallet located in a sports jacket hanging over a chair; c) stash the stolen money in his or her footwear (left foot); d) place the wallet back into the jacket; and e) report back to the laboratory assistant.

If the participant was in the "innocent uninformed" condition, the laboratory assistant asked him or her to go into the hall and read the instructions, which contained no information concerning the crime, walk down the hall to a particular location, wait, and report back to the laboratory assistant. The "innocent and informed" participants read material describing the crime with the same information given to the guilty subjects, but were instructed not to do the crime. Instead, they were to walk down the hall, wait at a particular location, and then report back to the laboratory assistant.

Once participants returned to the assistant, regardless of their condition of guilt or innocence, they were instructed to act as if they were innocent by co-operating with the interrogator and by answering truthfully

questions not specifically related to the crime. They were told to deny any questions relevant to the theft. This was the truth for innocent participants but was lying for guilty participants. It was stressed that the polygraph interrogator was unaware ("blind") of their actual condition so that judgment rested solely on their performance on the polygraph test. Five dollars was promised to each participant if his or her polygraph record was judged as innocent.

Only the participant and the assistant actually knew what condition he or she was in prior to and during the interview. The interrogator needed to know the order of questioning. This information was supplied through the assistant handing him the question sheet from the file when escorting the participant to the test room. The pretest interview was standardized. It consisted simply of reading the questions to participants. They were told to co-operate in general but to deny incriminating "crime-relevant" questions even if it meant "lying" for the guilty participants. The polygraph interrogator also showed the instrumentation to each participant. After the physiological measuring instruments were attached, one of the two polygraph examination question orders was administered with three repetitions. The following is the question list for Order 1. Critical or "crime-relevant" questions 4, 6, and 9 are before "control" questions 5, 7, and 10. Order 2 was created by reversing questions 4 and 5, 6 and 7, and 9 and 10.

1. Is your last name ____?
2. Are you afraid that I will ask you a question that was not reviewed with you?
3. Do you intend to answer each question truthfully?
4. Did you steal \$20?
5. Did you steal \$30?
6. Did you take the money out of a wallet?
7. Did you take the money out of a purse?
8. Is your first name ____?
9. Did you stash the money in your footwear?
10. Did you stash the money in your pocket?

After the examination, the assistant took participants and had them perform both recall and recognition tests concerning the theft. They received 50 cents per item recalled or recognized. They were assured that the interrogator would have no knowledge of any participant's condition or memory scores prior to marking the polygraph examination. They were also told that, following the entire data collection, copies of the educational component of the study could be obtained from the main Psychology Office.

Data Analysis

The major analyses involved 3 (Guilt Condition) \times 2 (Question Order) \times 2 (Gender) ANOVAs on scores from four physiological measures as dependent variables. Nominal significance was held at the .05 level for each analysis.

The dependent measures were derived through "blind" scoring of physiological recordings by comparing the relative magnitudes of responses in "control" and "crime-relevant" question pairs. The largest SRR amplitude in a 10-second period following each question was assessed with a ruler. If the response was larger to a "control" question than to its paired "crime-relevant" question, a positive one was assigned to that pair. If the response was larger to the paired "crime-relevant" question, a negative one was assigned. The linear distances of respiration curves for thoracic (THR) and abdominal (AB) recordings were measured with the use of an Alvin 1112 contour map wheel. The measurement period was for 10 seconds following the start of each question. Because suppression of respiration is associated with deception (Timm, 1982), the shortest distance was scored as plus or minus one depending upon whether it corresponded, respectively, to "control" or "crime-relevant" questions. BV variations were assessed with the use of a ruler in the 20-second period following a question. The largest upward excursion of the blood pressure/pulse envelope was identified and the amplitude of that upward excursion was measured. The largest such changes were assigned a +1 to a "control" question and a -1 to a "crime-relevant" question.

Classification of participants as guilty or innocent was done from the total of each individual's +/- score from each question pair for the four measures over the three repetitions of the three pair test. On the basis of a single measure, a participant's score could range from + to - 9. With the four measures together, scores ranged from + to - 36. Regardless of whether the range was for total scores or for single measure scores, individuals scoring +2 and above were classed as innocent whereas those scoring -2 and lower were classed as guilty.

The amplitudes of SRR scores for each of six "control/crime-relevant" questions were averaged across trials to create three pairs of six scores that were presented as the first, second or third pair of a trial. These scores were analyzed with an ANOVA. Condition, order, and gender were between-subject factors and question position and pair position were within-subject factors.

Results

Derived Scores

Condition effects occurred because of differences in SRR, $F(2, 108) = 8.2$, $\eta^2 = .13$, and BV, $F(2, 108) = 6.1$, $\eta^2 = .10$, measures but not with respiration measures. Duncan's Multiple Range Test found that innocent uninformed participants ($M = 0.25$) had average SRR scores more towards innocence than innocent informed participants ($M = -0.48$) who, in turn, scored as more innocent than guilty participants ($M = -2.48$). Duncan's Multiple Range Test on BV means found that both the innocent-uninformed ($M = 0.27$) and innocent and informed participants ($M = 0.73$) scored more towards innocence than the guilty participants ($M = -1.15$).

An order effect showed that derived SRR scores associated with "crime-relevant" questions in the first position (-2.73) were more in the guilt direction than those with the "crime-relevant" questions in the second position, + 0.93; $F(1, 108) = 4.17$, $\eta^2 = .04$. The same pattern of results held for BV. Scores associated with crime questions in the first position (-1.47) were more in the guilt direction than scores associated with "crime-relevant" scores in the second position, +1.37;

$F(1,108) = 3.86$, $\eta^2 = .03$. Again, respiration measures showed no differences.

Classifications

The classification of groups of participants into guilty, innocent, and inconclusive categories by various measures is shown in Table 1.

Hypothesis tests were done with inconclusive judgments excluded. When a

judgment of guilt or innocence was made, two measures (BV and SRR) plus a measure representing the total of all scores showed more guilty (73%, 86%, 82%) than innocent informed (36%, 48%, 50%) or innocent uninformed participants (42%, 46%, 48%) classed as guilty with the following significant chi square values: BV ($\chi^2 = 9.4$, $df = 2$) and SRR ($\chi^2 = 5.4$, $df = 2$) and total ($\chi^2 = 4.6$, $df = 2$). No differences were found for respiration measures.

Table 1
Numeric Classification Accuracy of Guilt, Innocence, and Inconclusive Judgments

Measure	Guilt			Actual Condition Innocent-Informed			Innocent		
	Judgment								
	G	I	Inc.	G	I	Inc.	G	I	Inc.
Total Score									
Order 1	16	2	2	14	1	5	13	4	3
Order 2	11	4	5	2	15	3	2	12	6
Abdominal									
Order 1	8	5	7	9	5	6	6	6	8
Order 2	7	3	10	3	8	9	4	8	8
Thoracic									
Order 1	4	3	13	7	5	8	8	2	10
Order 2	7	6	7	5	4	11	4	0	16
Blood Volume									
Order 1	14	2	4	7	2	11	6	1	13
Order 2	8	6	6	2	14	4	2	10	8
Skin Resistance Response									
Order 1	16	1	3	12	2	6	10	3	7
Order 2	8	3	9	1	12	7	2	11	7

Note. Order 1 = Question Presentation Order 1; Order 2 = Question Presentation Order 2. In the decked heads, "G" = Guilty; "I" = Innocent; "Inc." = Inconclusive.

When "crime-relevant" questions were positioned first, more participants were judged guilty across the same three measures (69%, 78%, 74%) than when "crime-relevant" questions were second (31%, 22%, 26%). The corresponding tests, respectively, were BV ($\chi^2 = 13.1$, $df = 1$), SRR ($\chi^2 = 14.9$, $df = 1$), and total ($\chi^2 = 13.5$, $df = 1$). No differences were found for respiration measures.

Chi square analyses showed that correct judgments depended on both the guilt condition and question position for BV ($\chi^2 = 14.9$, $df = 1$) and SRR scores ($\chi^2 = 13.08$, $df = 1$) and total scores ($\chi^2 = 13.50$, $df = 1$). Refer to Table 2.

Table 2
Overall Numbers of Participants Judged Correctly in Each Condition

Measure	Correct Guilty	Correct Innocent-Inf.	Correct Innocent	Total
Total Score				
Question Order 1	16	1	4	21
Question Order 2	11	15	12	38
Blood Volume				
Question Order 1	14	2	1	17
Question Order 2	8	14	10	32
Skin Resistance Response				
Question Order 1	16	2	3	21
Question Order 2	8	12	11	31

Note. "Innocent-Inf." = Innocent-Informed. "Question Order 1" = crime-relevant question first in pair; "Question Order 2" = control question first in pair.

Analyses of correct classifications for BV, SRR scores, and total scores over conditions and position shows that the "crime-relevant" questions in Position 1 resulted in 88%, 94%, and 89% correct classification rates, respectively, of guilty participants but only 19%, 19%, and 16% correct rates with innocent participants. In Position 2, the rates of correct classification were 57%, 73%, and 73% for guilty participants and 86%, 88%, and 87% for innocent participants.

Chi squares on the number of correct versus wrong classifications showed no measure resulted in classification accuracies above chance in the first position. In the second position, above chance classifications were found for blood volume ($\chi^2 = 8.6$, $df = 1$), SRR ($\chi^2 = 15.6$, $df = 1$), total scores ($\chi^2 = 18.3$, $df = 1$) but not for either of the respiration measures.

Memory Scores

ANOVAs conducted with recall and recognition memory scores as dependent variables and Gender, Condition, and Order as between-subject factors showed condition effects for both recall, $F(2, 108) = 70.12$, $\eta^2 = .56$, and recognition, $F(2, 108) = 78.21$, $\eta^2 = .59$. No significant effects were found for either Gender or Order. Subsequent post hoc analyses using the Tukey HSD method found

that innocent participants given no guilty information (M recall = 1.13, M recognition = 1.30) showed chance levels of memory, whereas participants given information in the innocent informed condition (M recall = 2.75, M recognition = 2.83) and guilty condition (M recall = 2.90, M recognition = 3.00) remembered that information.

Measured SRR Amplitude

An ANOVA on measures of SRR amplitudes found that males ($M = 8.8$) had lower magnitude responses than females ($M = 12.8$, $F(1, 108) = 6.86$, $\eta^2 = .06$). A serial effect, $F(2, 216) = 8.59$, $\eta^2 = .07$, found that responses declined from 11.8 for the first pair of questions, to 10.9 for the second pair, to 9.6 for the third pair. Questions in the first position of a pair ($M = 11.4$) were larger than those in the second position ($M = 10.1$; $F(1, 216) = 14.3$, $\eta^2 = .12$). A Condition by Order interaction, $F(2, 108) = 3.79$, $\eta^2 = .07$, followed by Duncan's multiple comparison test showed that the response magnitudes to the average of "control" and "crime-relevant" questions were larger in the innocent uninformed group when "crime-relevant" questions were in the first position than for any other group and position except for guilty participants responding to the average of "control" and "crime-relevant" questions in the second position. See Table 3.

Table 3

Means and Standard Deviations for SRR Amplitudes in Order of Question Presentation Across Conditions

	Condition 1	SD	Condition 2	SD	Condition 3	SD
Order 1	15.1	14.2	8.7	5.6	8.6	7.1
Order 2	9.4	5.7	9.6	6.2	13.2	10.1

Discussion

Support for the main hypothesis was found. The relative size of physiological responding in "control" and "crime-relevant" question pairs is determined by order or the serial position of questions and guilt condition. The first question in a pair evokes the larger response and responses diminish over successive pairs. This finding held even in the innocent informed group. In this group, participants knew the item was relevant to the crime. The important exception to this general finding occurs when participants are deceptive to the second question. Items in the second position are evocative enough to generate responses that exceed ORs when participants are deceptive to those items. Knowledge alone was not enough. When the "crime-relevant" question was first, participants, regardless of their assigned condition, responded as if they were guilty. Derived scores and classifications based on the composite of all scores, SRR scores, blood volume scores, and SRR amplitudes were the source of these results.

Classification of participants showed that there is little chance of correctly classifying innocent participants as innocent when "crime-relevant" questions are in the first position. When "crime-relevant" questions were in the second position, conclusive classifications were accurate for both guilty and innocent participants. The numbers in the results are similar to those reported in a review of standard CQT's by Ben-Shakhar and Furedy (1990). In that review of nine studies, they found 80% of the guilty and 63% of the innocent subjects were classified correctly. They reviewed an additional nine field studies of the CQT and found rates of 84% and 72% correct classifications for the guilty and innocent.

The study, the analyses, and the psychological conditions for the participants created some complexities that we did not fully predict. Empirically, the data conformed to our expectations at the classification level but less so at other levels. For example, there were two ways that clear support for the main hypothesis could be found in the derived score analysis. The way hypothesized was through a condition by order interaction. We thought large responses would occur to "crime-relevant" questions in the first position for all participants; regardless of condition, but would occur to "crime-relevant" questions only for those in the guilty condition when the "crime-relevant" questions were in the second position. Empirically, however, only a condition effect was found. Support for the hypothesis with classifications was because the cutoff point between positive and negative scores played a role that is reflected in the classification data. When "crime-relevant" questions were in the first position, all participants collectively scored as guilty (below zero), with the guilty scoring as most guilty (most negative mean score). When "crime-relevant" questions were in the second position, scores across all groups became less negative to the point that scores were positive in the innocent groups but still remained negative in the guilty group. The three groups remained parallel for the two order conditions (thus no interaction) but were altered in relationship to zero and the cutoff points for guilt/innocence judgments.

The analysis of SRR amplitudes revealed that response magnitudes generally followed a course of habituation expected through the serial position of items. Initial items evoked large responses and subsequent items evoked smaller responses. The exception was in the guilty condition where items to be

deceptive on were in the second position. The results are relatively straightforward but again there are complications. The best simple result of the analysis would have involved a three-way Condition, Order, Question type interaction with guilty participants having large magnitude responses when the "crime-relevant" question was in the second position. Empirically, question type was not a factor. Instead, a two-way condition by order interaction was found since a) initial questions regardless of crime relevance were very large for innocent uninformed participants and b) knowledge of "crime-relevant" questions was not a strong generator of responses when participants were not deceptive.

To explain the results, we speculate that although the OR and habituation are strong factors in detection of deception paradigms, the context in which these processes take place is also very important. We think the structure of the test makes the most sense for informed participants. The pretest interview and the test itself consisted of reading questions that guilty or innocent informed participants see as completely obvious. Everyone is being asked questions on which all will tell the truth paired with questions on which guilty participants will lie. On a face validity basis, it seems straightforward and fair. The innocent tell the truth and from their perspective that can be appropriately judged. They realize, also, that guilty participants will be lying and that too should be very evident on a question pairs basis. The uninformed group could feel a little less certain. The test is less transparent to them because they do not know the "crime-relevant" items. These speculations could be examined in future studies. For example, if curiosity about the crime were a factor for the uninformed group, they could be asked to identify which items they thought were associated with the crime.

Although the manipulations and modifications in this study are of such a basic nature as to be fundamentally obvious, the results make a contribution to aspects of a "heated debate" (Ben-Shakhar & Furedy, 1990; Lykken, 1981; Raskin, 1978) in lie detection. For these authors, the focus of the debate has been on the content of questions. This study shifts the focus to position. We

removed the questionable features of "control" questions. The comparability of the two types of questions, save for crime relevance, was equated. With such changes, position becomes a major determinant of responsiveness. Questions in the first position evoke larger responses regardless of whether they are "control" or "crime-relevant" questions. Questions in the second position evoked larger responses only if participants were deceptive. Knowledge without deception to questions in the second position did not result in larger responses. In the experimental laboratory context, the OR effect alone provides strong protection against erroneous classification for innocent participants when the "control" question is first.

The implications of these findings are considerable. Since, in this present study, the OR alone accounts for the reactions to "control" questions, the reputed attributes deemed necessary for standard "control" questions may not be necessary as long as "control" questions are in the first position. In essence, disputes over question attributes (Lykken, 1981) and the use of the word "control" in "the true scientific sense" may have been misguided. All that may be necessary is a foil that evokes an OR and initiates a habituation process for the subsequent "crime-relevant" question. If our results can be replicated by other researchers and if they generalize to field conditions, formulation of adequate "control" questions could be simply prescribed. The necessary rule to follow would be to have "control" questions that change the topic or category from prior questions but that would be on the same topic as subsequent "crime-relevant" questions. In other words, the new topic must be introduced for an initial OR but for the second question either habituation or deception is the determinant of whether the response is smaller or larger. For example, if a prior set of questions addressed the amount of money stolen, the next set could address the weapon used or the escape vehicle, etc. One great strength to this approach is that an interview prior to the administration of the test could be standardized to "These are the questions. If you are innocent, you will be telling the truth on all questions. If you are guilty, you will be telling the truth on some questions but not others."

Our results suggest a potential for abuse of the CQT. A simple way to make a suspect appear guilty would be to place "crime-relevant" questions first in the set. The examiner naive or insensitive to order effects and the OR could cause great consternation by finding innocent people guilty.

Although we have described the sufficient conditions to evoke responses to "control" questions, it is probable that these responses could be enhanced with additional content. This idea comes from findings with guilty participants. For them, the magnitudes of physiological responses to "crime-relevant" questions in the second position exceeded the magnitudes of OR responses to "control" questions. This occurred because the content for guilty participants was of unique importance. The guilty participants had committed actions involved with the questions and had to lie to "crime-relevant" questions. It was just not simple knowledge that evoked responding since innocent participants with knowledge were not any more reactive than uninformed participants. The knowledge has to be relevant, of importance, and associated with a lie. It is not a stretch to imagine designing highly evocative "control" questions based on content. With this understanding, it is now possible to do studies that assess the effect of adding in "control" question features alleged to be important and to examine them in a systematic, scientific way. The effects of lying, importance, relevance, degree of incrimination, ambiguity, and position could all be looked at. For example, as is the case in many real criminal investigations, many suspects innocent of a particular crime under investigation are hiding aspects of a criminal history. Thus, they may well be responsive to standard "control" questions and the CQT would generally be effective and possibly even if "control" questions were in the second position. It still remains an empirical question but one that now can be examined systematically rather than be argued about on some assumptions by one side or the other.

Memory scores showed a condition effect for both recall and recognition. Groups with information (innocent-informed and the guilty groups) had almost complete recall and recognition memory of the crime material. While not different from each other, they

differed from the innocent uninformed group who had memory scores at chance levels. Even though the guilty and innocent informed groups had similar scores, only guilty participants were detected as guilty when "crime-relevant" questions were in the second position. This finding agrees with those of Waid, Orne, Cook, and Orne (1978), Waid, Orne, and Orne (1981), and Iacono, Boisvenu, and Fleming (1984) and suggests that knowledge, while a necessary condition for detection, is not a sufficient factor for detection. Psychological factors associated with some or all of context, roles, deception, and meaningfulness play a role in physiological responsiveness.

A caveat applies when considering this or any other laboratory study where the success of the test depends upon memory for details. In typical studies (e.g., Bradley et al., 1996), participants are required to remember what researchers believe is pertinent crime-relevant information. Participants, often students, typically read over and study the important material (e.g., Waid et al., 1981). In real life, however, items selected by investigators may not actually be the type of information that suspects remember. Thus our success in detecting guilty suspects with "crime-relevant" questions in the second position may be difficult to duplicate in the field if criminals forget the basic information about their crime.

There are, of course, important caveats regarding innocent suspects. Laboratory studies do not approach the emotional and tension levels associated with field situations. It is completely possible that knowledgeable suspects in the field may have reactions to "crime-relevant" questions that exceed those of the ORs to "control" questions. Thus, more understanding of this test format is necessary before considering any application. The advantage of the test format is that response-generative elements can now be studied empirically rather than just be discussed on assumption-laden logical grounds.

In summary, the "control" questions used in this study are unlike normal "control" questions. There is no ambiguity, probable lie assumption, or "emotional" content (Bradley et al., 1996). That is, there is no pairing of a

"crime-relevant" question such as, "Did you steal the money from the drawer?" with an emotionally laden control question such as "Did you ever steal anything of value?" The present test is more like a GKT, but with only one buffer question, such as "Did you steal \$20?" or "Did you steal \$30?"

The results of this study support the contention by Bradley et al. (1996) that "control" questions do not have to have their alleged characteristics. They must, however, in the same manner as distractor items in the GKT, be plausibly related to the crime. This requirement may even be questionable, since informed participants were aware that the "control" questions were not related to the crime but still responded more to those questions, due to the OR effect, if they were presented in the first order.

The CQT with modified "control" questions addresses several issues surrounding the CQT. First, the questions involved in the test may be inherently standardized. That is, the "crime-relevant" questions would be on an item related to the crime and, in the same format and style, the "control" questions would be on a matched item but one that simply is not involved in the crime.

Secondly, at the present time CQT interviews are not well standardized (Iacono & Patrick, 1988) although standardization is required for any psychological test (Anastasi, 1988). Our modified CQT eliminates the need for a complicated interview and it may be readily standardized. There would be no need to draw particular attention to "control" question features or to match "control" questions to the suspect. A rather brief interview can convey the only feature that needs to be made clear. That feature is that innocent suspects are telling the truth to both types of questions whereas guilty suspects are lying on one type.

Third, a transparency problem (Ben-Shakhar & Furedy, 1990) is dealt with directly. This problem arises because both

innocent and guilty subjects know that the "crime-relevant" questions are the most important to "pass successfully" to avoid a judgment of guilt. Due to various emotional reactions, it has been argued (Lykken, 1981) that innocent suspects will produce responses similar to those of a guilty suspect on "crime-relevant" questions. Empirically, in the present study and in the previous study by Bradley et al. (1996), informed innocent participants simply did not respond to the incriminating "crime-relevant" questions in the second position. Further, if innocent suspects are uninformed about crime details then neither of the questions in each pairing is transparent to the innocent subjects. Therefore, those who are innocent and uninformed truthfully answer "no" to both questions with no knowledge of which is which, whereas the guilty will answer truthfully to one question and lie in responding to the other item. The ultimate consequence is that, with the "crime-relevant" question second in a pair, innocent suspects will respond to the initial "control" question whereas guilty suspects will respond more to the "crime-relevant" question.

A fourth point is that our modifications have created an alternate form of GKT with all of its inherent strengths if innocent suspects are unaware of information. Of course, it retains the glaring weakness of the GKT with guilty suspects but that weakness may be somewhat mitigated. The obvious problem for the GKT is that suspects may not remember the information chosen for the investigation. Our version uses few items. Thus, the investigator can choose the most obvious information in the hopes that suspects retain that material.

The fifth and most important point from our perspective as scientists is that this study brings up the issue of scientific clarity. The past 50 years of discussion of the "It works well" to "It does not work well" variety will extend to another 50 years unless researchers (we in this case) begin to identify systematically with precision the elements in questions that are effective.

References

- Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.
- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan Publishing Company.
- Ben-Shakhar, G., Asher, T., Poznansky-Levy, A., Asherwitz, R., & Lieblich, I. (1989). Stimulus novelty and significance as determinants of electrodermal responsivity: The serial position effect. *Psychophysiology*, 26, 29-38.
- Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception*. New York: Springer-Verlag.
- Ben-Shakhar, G., & Lieblich, I. (1982). The dichotomization theory for differential autonomic responsivity reconsidered. *Psychophysiology*, 19, 277-281.
- Bradley, M. T., MacLaren, V. V., & Black, M. E. (1996). A true control polygraph control question test. *Perceptual and Motor Skills*, 83, 755-762.
- Bradley, M. T., & Rettinger, J. (1992). Awareness of crime-relevant information and the Guilty Knowledge Test. *Journal of Applied Psychology*, 77, 55-59.
- Honts, C. R. (1993). A review of "Theories and applications in the detection of deception." *Psychophysiology*, 30, 317-319.
- Iacono, W. G., Boisvenu, G. A., & Fleming, J. A. (1984). Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology*, 69, 289-299.
- Iacono, W. G., & Patrick, C. J. (1988). Assessing deception: Polygraph techniques. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 205-233). New York: The Guilford Press.
- 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.
- Lykken, D. T. (1959). The SRR 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.
- Raskin, D. C. (1978). Scientific assessment of the accuracy of the detection of deception. *Psychophysiology*, 15, 344-359.
- Reid, J. E., & Inbau, F. E. (1977). *Truth and detection, the polygraph ("lie detector") technique* (2nd ed.). Baltimore, MD: Williams and Wilkins.
- Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing. *American Psychologist*, 3, 355-366.
- Timm, H. W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10, 47-51.

- 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 information. *Journal of Applied Psychology*, 63, 728-733.
- Waid, W. M., Orne, E. C., & Orne, M. T., (1981). Selective memory for social information, alertness, and physiological arousal in the detection of deception. *Journal of Applied Psychology*, 66, 224-232.

What does the Photoplethysmograph Indicate?¹

L. A. Geddes²

Plethysmography is concerned with recording the volume of a body segment. The name is derived from the Greek word meaning fullness. Therefore, it would appear that a plethysmographic recording ought to provide a single type of physiological information, namely whether there is more or less blood in the segment to which the plethysmograph has been applied. However, in a practical case, the situation is not so simple. It will be the object of this brief report to describe what the standard, available plethysmographs record and what physiological information is contained in such recordings.

Principles of Operation of Photoelectric Plethysmographs

Basically there are two types of photoelectric plethysmographs; one type

operates via light transmission, the other employs light reflection. Figure 1 illustrates both types. Usually visible light is employed for measurement. A change in the volume of blood in the transmission or reflecting path will therefore alter the amount of light presented to the photoelectric detector.

It would be simple enough to state what information is contained in the plethysmographic record if the method were applied in a straightforward manner, for movement of the baseline would indicate an increase or decrease in the volume of the region of the body to which the plethysmograph has been applied. Small pulsatile changes in volume would ride on any shift in baseline which indicates a change in blood volume below the photoplethysmograph. Such a recording system would

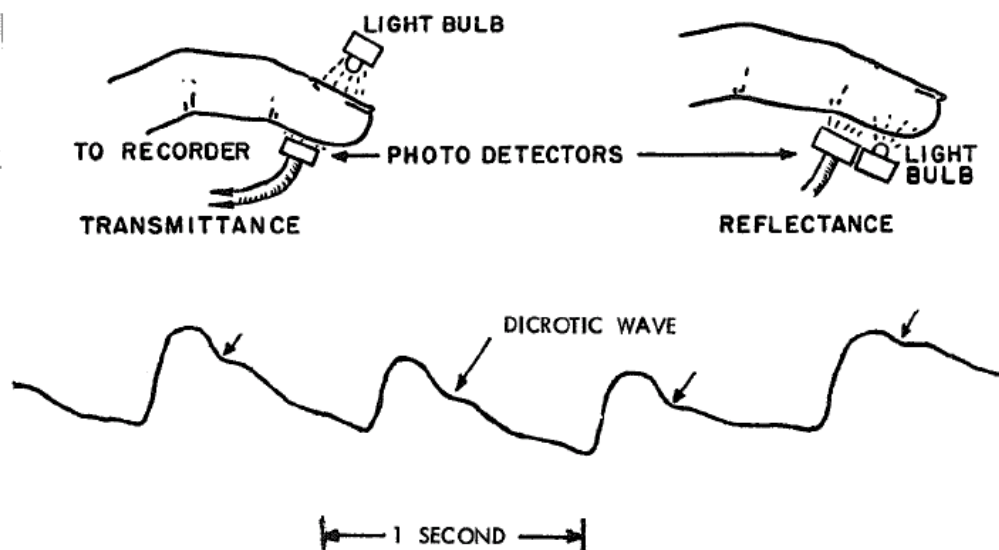


Figure 1. The two types of photoplethysmograph and a typical record. In one type the amount of transmitted light is measured and in the other, the amount of back-scattered reflected light is measured.

¹ This article originally appeared in *Polygraph* in 1974, volume 3(2), pages 167-177. It was reprinted at the request of a member of the American Polygraph Association member.

² Director Division of Biomedical Engineering Department of Physiology College of Medicine Houston, Texas 77025.

be designated a direct-coupled or dc system. However, with such a system, the display of the volume shifts that occur when tiny blood vessels contract or relax, result in the appearance of only very tiny pulsatile oscillations appearing on the record. Therefore, to better visualize the pulsatile volume changes the overall amplification of the recording system is increased and direct-coupled recording is not employed; instead, capacity coupled recording is used to obtain a stable baseline. Therefore, the information on true volume shifts in the body segment is lost, although moderately rapid changes are detectable but not quantifiable.

Figure 2 presents schematically the meaning of the statements made in the previous paragraph. An increase in the amount of blood in the segment is caused to occur suddenly as shown in Figure 2A. Note that the capacitively-coupled plethysmograph (Figure detects only the change in the volume, rather than demonstrating that a sustained increase in volume had occurred. When the volume of blood in the segment is suddenly reduced, only the change is recorded and the recording returns slowly to the original baseline as in the previous case. The time taken for the recording to fall to 37% of the peak amplitude of the change is called the time constant which is made short enough to eliminate slow baseline variations which reflect true volume changes. In all commercially available models the time

constant is made long enough to allow recording of the pulsatile changes in blood volume. Often, however, the time constant is too short to permit display of the slow respiratory volume changes.

It is possible to create a photoplethysmograph system which can display total volume change in a body segment. However, when this is done, it becomes extremely difficult to keep the baseline of the tracing centered on the record because slight displacement of the photoplethysmograph, and small volume changes in the underlying tissues, will produce large displacements in the baseline of the recording. In addition, the pulsatile changes in amplitude would be quite small. However, with capacitive coupling, volume shifts do not show up on the baseline and the recording appears much more stable and it is possible to increase the sensitivity to make the pulsatile changes clearly visible. If a sustained volume increase occurs with a rapid onset, the baseline will be deflected during the change and will return to its former level, as sketched in Figure 2B, despite the fact that the body segment has a new volume. If the volume of the segment decreases quickly, there will be a transient deflection in the baseline in the opposite direction. Obviously very slow changes, and sustained changes in volume of the segment, will not be detected by capacitively-coupled plethysmographs.

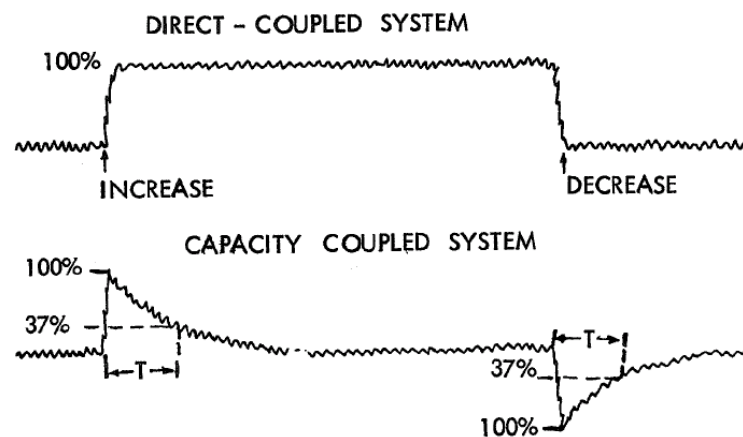


Figure 2. Schematic representation of direct-coupled (A) and capacitively-coupled photoplethysmographic systems (B). A sudden increase and latter sudden decrease in blood volumes is represented in A. With conventional capacitatively-coupled photoplethysmographs, only the changes are displayed, as shown in B. The time taken for the recording to fall from 100% of the change to 37% is called the time constant (T) and is measured in seconds.

The importance of an adequately long constant is demonstrated in Figure 3, which shows a typical recording made on a relaxed subject as the time constant was increased from 0.15 to 2 seconds. Note as the time constant was increased, the recorded pulsatile amplitude increased and, in addition, the respiratory induced volume changes become recordable.

Because capacitive coupling is used for convenience in recording, one might well ask if there are changes in the amplitude and contour of the photoelectric pulse which indicate that changes in blood volume of the segment have occurred. There often are, but the changes are small and difficult to recognize. Two types of change can occur; one relates to the overall amplitude, which is decreased with vasoconstriction; the other is a change in the dicrotic wave (Figure 1) which often becomes less pronounced. However, in many subjects, the dicrotic wave is not identifiable.

Despite the fact that the capacitively coupled plethysmograph only indicates transient changes in segmental volume, it does show heart rate. In addition, if the time constant of the capacitive coupling is long enough, respiratory variations can be seen varying the baseline and the amplitude of the recording as shown in Figure 3.

Physiological Responses Recordable with the Photoplethysmograph

A large number of vital body functions (blood pressure, heart rate, temperature, water balance, secretion of glands, digestion, elimination, etc.) are controlled by the autonomic nervous system, which functions virtually unnoticed. The autonomic nervous systems [sic] consists of two parts, the sympathetic and parasympathetic. Both parts participate in regulation of the functions just identified. Although the autonomic nervous system operates by itself, its activity is modulated by changes in the mental state of a subject. Each person has his own pattern of response to an alerting or threatening stimulus, and this mental response alters the activity of the system.

Activation of the sympathetic division of the autonomic nervous system causes dilation

of the pupils, a tendency toward dryness in the mouth, an increase in heart rate and blood pressure, vasoconstriction in some vascular beds, (especially the skin), cessation of the activity of the gastrointestinal tract and the secretion of sweat. Activation of the parasympathetic division of the autonomic nervous system causes constriction of the pupils, salivation, slowing of the heart rate, increased activity of the gastrointestinal tract and evacuation of the bladder and bowel, if voluntary control does not supervene. The sympathetic nervous system tends to produce all of its effects; the parasympathetic is more discrete and capable of more variety in the type and degree of response.

It is an interesting fact that when a person is presented with an alerting or threatening stimulus, he can only conceal certain normally visible responses; autonomic nervous system responses cannot be entirely suppressed voluntarily. Thus a broad spectrum of physiological events is available to indicate the response to an alerting or stressful stimulus. It is the autonomic response, along with respiration, that polygraph examiners record during interviews.

It is now useful to relate the information provided by the photoplethysmograph to the physiological events (skin resistance, breathing, heart rate and blood pressure) recorded by polygraph examiners. In doing so, it is important to recognize that the type of response to an alerting or threatening stimulus is highly individualized, a fact that is well known to examiners. Nonetheless, it is useful to investigate the possible relation of the photoplethysmogram to the respiratory, cardiac and skin resistance channels. In practice, whether it turns out to be so, or not, the changes seen in the photoplethysmogram should be related to those in the cardiac channel. A change in heart rate will, of course, show up in the photoplethysmographic record. Whether a change in blood pressure is indicated cannot be stated with certainty. Blood pressure is increased by vasoconstriction, but blood vessels in a variety of beds can be constricted to accomplish this response. In all probability, in some subjects, the vascular bed seen by the photoplethysmograph will constrict, and this event will be revealed by a transient movement in the baseline and a decrease in

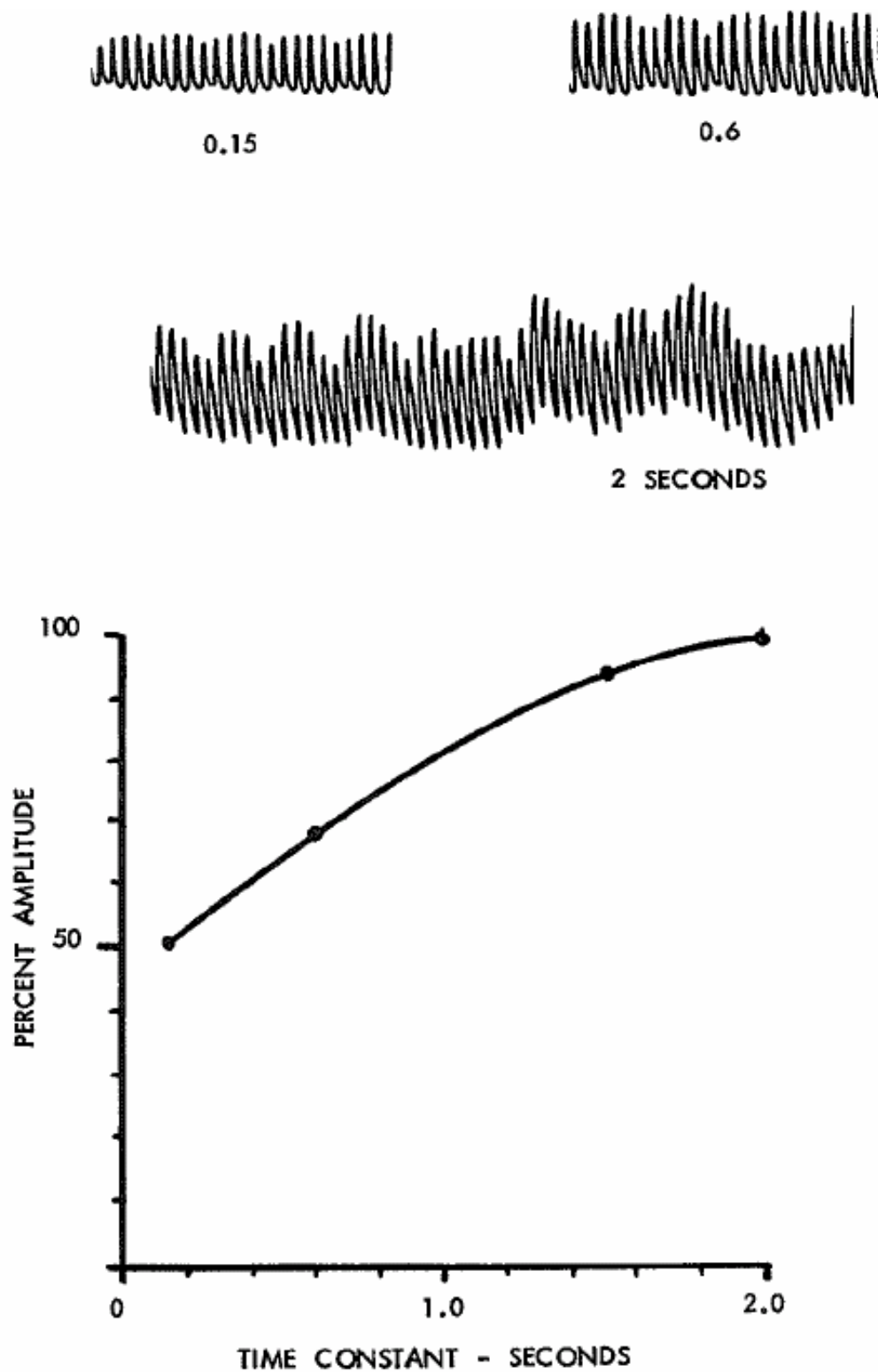


Figure 3. Changes in the photoplethysmogram produced by varying the time constant from 0.15 to 2 seconds. Note that as the time constant is increased, the pulsatile amplitude becomes larger and the slow, respiratory changes start to appear. The graph below [sic] shows that a time constant of about 2 seconds is adequate for displaying respiration and the pulse.

overall amplitude of the pulse height, along with a diminution in the size of the dicrotic wave, if previously present. Since a skin resistance change, like an increase in blood pressure, is produced by an increased outflow of the sympathetic nervous system, a GSR should accompany a change in the plethysmogram. If the time constant of the photoplethysmographic channel is adequately long, the vasoconstrictive event may be signalled quite well by a transient shift in the baseline. The use of an adequately long time constant will also favor reproduction of respiratory variations in blood flow. Unfortunately, no exhaustive studies been carried out to date to identify the most appropriate time constant for the photoplethysmograph. With most of the available units, the time constant has been chosen only long enough to reproduce the pulse wave and to provide a baseline that need not be continually recentered. Such a situation may result in missing important respiratory-induced volume changes in the segment seen by the plethysmograph. What all of this means is, that although it is easy to make and use a plethysmograph, the information that it will produce depends on the type of circuit used with it and the subject's type of response. Because of ease of application and the fact that it can indicate cardiovascular events, there is need to conduct serious studies, first with direct-coupled plethysmographs to examine the true nature of the changes in segment volume encountered in polygraphic examinations, and then to discover whether the capacitively-coupled photoelectric plethysmograph can indicate them.

In the design of a photoplethysmograph, great care must be used to guarantee that the light source does not produce enough local heating and alter the degree of vasodilation or vasoconstriction that existed before the device was applied. To minimize this effect, many instruments use either a small, low wattage bulb and operate it below its rated voltage. Often a light-emitting diode (LED) is used which emits "cold" colored light in a narrow band. However, unless care is taken, the heat produced by the LED may also alter the local circulation. It is an interesting fact that a little heat produces a slight degree of vasodilation and provides a large amplitude pulsatile signal

from the photoplethysmograph. Just how much heating is permissible to obtain the most useful information for poly-graphic examination is not known as yet.

Despite the lack of adequate design information for photoplethysmographs to be used in polygraphic examination, it is possible to use some existing models profitably. For example, the responses to two different types of stimuli are shown in Figure 4. A reflectance type photoplethysmograph was applied to the tip of the second finger of the left hand. The overall time constant was 2 seconds. In Figure 4A, the subject was relaxing with his eyes closed and respiration can be identified as slow variations in the amplitude and baseline of the recording. The subject was instructed to inhale deeply and then exhale. Note the change in amplitude and shift in baseline of the record following the breath. Note also the increase and decrease in heart rate.

In Figure 4B, the subject was relaxed with his eyes closed, and the operator delivered an alerting stimulus by clapping his hands near the subject's ear. Note the transient decrease in amplitude and shift in the baseline of the plethysmographic record. On this occasion there was virtually no heartrate change.

From the foregoing, it can be seen that one of the factors of major importance appreciated with the photoplethysmograph is the time constant used with it. An adequately long time constant is necessary to reproduce the pulse accurately and to display respiratory variations. With a time constant of 2 seconds, recordings such as those shown in Figure 4 can be obtained and investigated for their value in polygraphic examinations.

References for further reading

Brown, C. *Methods in Psychophysiology*. Baltimore, Md., 1967. Williams & Wilkins Company.

Venables, P. H., & Martin, T. A. *Manual of Psychophysiological Methods*. Amsterdam, 1967, North Holland Publishing Co.

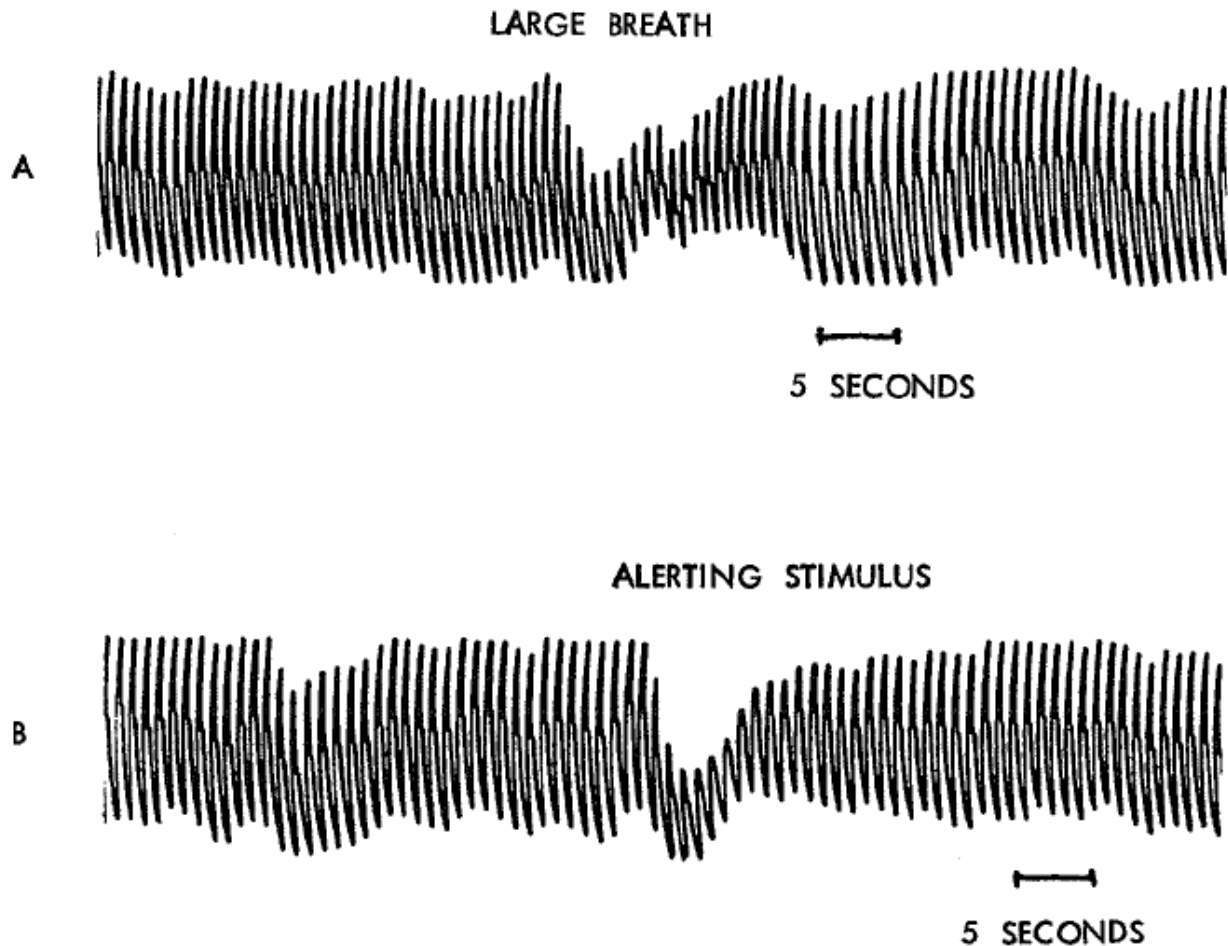


Figure 4. Photoplethysmograms recorded from a relaxed subject using a time constant of 2 seconds. In A, the subject was asked to take a deep breath; note the decrease in pulsatile amplitude, the shift in the baseline of the recording and the transient increase in heart rate. In B, the subject was pre-sented with an alerting stimulus (a loud hand-clap); note the transient decrease in pulsatile amplitude and shift in the baseline of the recording. In this case, no change in heart rate occurred.

Postscript: The Multigraph and Emotional Stress Monitor polygraph instruments produced by Stoelting Company offer three modes of plethysmograph operation. There is a D.C. coupled mode ("manual"), an A.C. mode with a time constant of 1.5 seconds ("Auto 1") and an A.C. mode with a time constant of second ("Auto 2". (Ed.)

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.