EUROPEAN POLYGRAPH

PUBLISHED SEMI-ANNUALLY

2022 VOLUME 16 NUMBER 2 (56)

ANDRZEJ FRYCZ MODRZEWSKI KRAKOW UNIVERSITY

EUROPEAN POLYGRAPH

PUBLISHED SEMI-ANNUALLY

2022 VOLUME 16 NUMBER 2 (56)

ANDRZEJ FRYCZ MODRZEWSKI KRAKOW UNIVERSITY

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
Journal of Andrzej Frycz Modrzewski Krakow University

European Polygraph is an international journal devoted to the publication of original investigations, observations, scholarly inquiries, and book reviews on the subject of polygraph examinations. These include jurisprudence, forensic sciences, psychology, forensic psychology, psychophysiology, psychopathology, and other aspects of polygraph examinations.
Opinions expressed in articles and book reviews published in *European Polygraph* solely reflect the experience and expertise of their respective Authors and Reviewers. Their publication does not imply any form of endorsement on behalf the Editors and Publishers who provide space for the presentation of, often contradictory, views and discussion of positions.

Office
ul. Herlinga-Grudzińskiego 1C;
30-705 Kraków (Poland)
e-mail: aszuba@afm.edu.pl
www.polygraph.pl

e-ISSN 2380-0550
ISSN 1898-5238

European Polygraph offers open access to all its issues:
http://www.polygraph.pl
https://sciendo.com/journal/EP
The original of *European Polygraph*
is its electronic version.

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

Contents

Articles

Book review

Articles

sciendo

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

DOI: 10.2478/EP-2022-0008

Thoughts on the Inconclusive Zone in Comparison Question Test (CQT)

Avital Ginton[1]
AV-GN, polygraph, applications & research[2] , Tel-Aviv, Israel

Abstract

The Inconclusive Zone, in one form or another, is an integral part of almost any data analysis method of CQT. It represents the existence of built-in uncertainties in the test and indicates how the test, in general, and the examiners, in particular, bear this limitation in practice. Unfortunately, it is so common that most examiners tend not to spend too much time and attention wondering about its meaning. The following are some reflections on the concept of an Inconclusive Zone, and its actual use, including some resulting recommendations. That includes internal aspects of the test, such as the relation between the extent of using multiple comparison points between relevant and comparison questions and the optimal Inconclusive

[1] Avital Ginton, Ph.D., a psychologist, past head of the Israel Police Polygraph section (1977–1994), avital.ginton@gmail.com.
[2] AV-GN, polygraph, applications & research is a private independent institute. The author also teaches part time in Tel Aviv University – The Buchmann Faculty of Law (c/o Prof. S. Shoham), Israel. Email: ginton@zahav.net.il Alternative Email: avital.ginton@gmail.com

Zone size. External aspects affecting the Inconclusive Zone's size and shape are also stressed when looking at the CQT through a prism of aiding decision-making rather than a mere means to sort truth from lies.

Key words: Comparison Question Test, CQT, Inconclusive Zone, Polygraph, True score theory, Pay-off matrix in polygraph decisions

Assuming that polygraph subjects know whether they are telling the truth or lying to the relevant questions, it is clear that an inconclusive result does not reflect their state of mind regarding the truthfulness of their answers to the relevant questions. Rather, it reflects the Comparison Question Test (CQT) method (Krapohl & Shaw, 2015) or its execution failure to identify the subjects' state of mind in that respect. More precisely, it is a failure to reach a definite conclusion about that, whether true or false.

In the CQT, the primary rule in deciding whether the examinee is found to be deceptive or truthful to the relevant questions states that if the magnitude of the response to the Relevant Questions is found to be stronger than to the Comparison questions, namely, **R >C** – the subjects are deemed Deceptive (DI) and if the opposite is detected, i.e., **C >R,** they are considered truthful (NDI). Thus, in the CQT, the indicator for lying or telling the truth to the relevant questions is the difference between the magnitude of physiological responses to the two categories of questions. The rationale behind these decision rules is that the response magnitude is highly correlated with the amount of concern from the questions. And while deceptive subjects are mostly concerned about the relevant questions, the opposite is found among truthful examinees (Vrij, 2008: 302)[3].

However, due to the ongoing fluctuations in the states of mind, cognitively and emotionally, and in the body's physiological activity, any two measuring of responses even to the same objective stimuli (i.e., questions) may subjectively be responses to somewhat different inputs that induce different output intensities (e.g., the same question might be detected as more or less intimidating in different repetitions)[4].

[3] This rationale has been critically questioned by many scientists in the academic world (Vrij, 2008: 304–311) but the present paper is not the place to argue on that.

[4] Physiological response to psychological stimulus (e.g., a question) is influenced by the psychological state of the individual at the moment when the stimulus is "thrown" at them. How focused they are? How much are they self-absorbed or lost in thought at that moment? To what degree they are aroused or excited at that particular minute? And so forth. All these are matters that by nature contain some sort of lability resulting in mental baseline fluctuations which modify the way

These fluctuations are considered background noise that obstructs the detection of the pertinent signal. One characteristic of the noise is its masking quality that interferes with identifying the signal. But once we manage to detect what seems to be a signal, we still should wonder whether this detected signal is enough to conclude about the truthfulness of the examinee regarding the relevant questions.

The theoretical rationale of CQT states very clearly the expected direction of the differences between R and C questions; however, it says nothing about the size of these expected differences in the magnitude of reactions.

Why is it then that we are looking for a certain minimum in the magnitude of the difference before rendering a decision as manifested in applying an Inconclusive Zone? Somehow we assume, probably justifiably, that slight differences due to incidental noise might

frequently occur, as opposed to big differences, which are very rare to occur due to random fluctuations per-se. Hence we want to see a big enough difference before attaching any significance to it. While that might seem quite logical, one should keep in mind that to avoid relying on an observed difference that occurred just by random fluctuations or other kinds of non-systematic noise, a consistency factor is much more important than the size of the difference (Ginton, 2013).

To reduce the subjectivity of estimating differences in response magnitudes, a technique attributed to Cleve Backster (1962, 1963), who termed it Numerical Scoring, was developed. For those unfamiliar with the numerical scoring technique, the following is its skeletal description: It is based on comparing the detected physiological responses to a Relevant question each time it is presented with those of an adjunct Comparison question. This is done per each physiological measure, which usually consists of Respiratory, Cardiovascular, and Electrodermal activities. Each comparison is considered a separate comparison point, and the difference in the response magnitude in each point gets a number

---

the individual experiences the stimuli. Therefore, even if we objectively throw the same stimulus (a specific question) at the subject twice, they may experience them differently (as long as we assume fluctuations in his momentary mental state) and react with different physiological responses. To that, we should add the fact that the mere body activity is constantly in flux, and the physiological response is always built on top of the physiological state of the responder at the said moment. The bottom line is that there is a high probability that two stimuli that are objectively considered equals create different reactions. Because of this potential variation, it is necessary to repeat the stimuli several times to pick the commonality and reject the momentary differences.

indicating the difference in the response strength between the compared questions in that particular point. If there is no substantial difference, the number is Zero; if the Relevant response is stronger than the Comparison, it gets a Negative number, and a Positive number is given to a comparison point, indicating a more robust response to the Comparison question relative to the Relevant. Originally the scoring technique applied a seven points scale from −3 to +3, indicating the degree of difference found in each comparison point and its direction. However, later a three-point scale indicating only the direction of the difference was also adopted, i.e., −1, 0, +1. The numbers from all the comparison points, including repetitions of questions, add to a total score across repetitions of each relevant question and, when applicable, across all the relevant questions together. Based on prior decided positive and negative cut scores[5], the examiner could determine whether the difference between the Response's magnitude to R and C questions is big enough to make a call. A Negative final score that reaches the cutpoint means a Deceptive examinee, and a Positive score indicates a Truthful examinee. The range between the two cut scores (the negative and the positive) is an Inconclusive Zone, meaning that getting a final score in this range is not enough to make a call. Thus, the moment of birth of the Inconclusive Zone was when the decision rules in numerical scoring set two separate cut scores, one as the threshold for a DI call and the other as the threshold for an NDI ruling. This does not mean that with smaller differences, the rate of correct decisions that would have been made if these were determined by the mere direction of the difference is similar to flipping a coin. It should still be above random success; However, the estimated error rate may be too high to be used as a practical method of distinguishing lies from truths. For instance, Ginton (2012) showed that relying on a decision criterion that did not leave any undecided test, which practically means not having an Inconclusive Zone, resulted in an error rate of twenty percent. Is that still good enough to be used for decision-making, or is it too costly? That is a matter of policy.

The Inconclusive Zone has been an integral part of any numerical scoring technique and, with some needed adaptation, also part of many computerized data analysis algorithms. Hence, discussing various aspects of this practice seems essential to the polygraph profession, as presented in the following reflections on that issue.

[5] Based on theoretical reasons, on empirical research or just arbitrarily.

The Inconclusive Zone and the True Score Theory

As already mentioned, the CQT is based on the assumption that deceptive examinees are more concerned about the relevant questions than the comparison questions. In contrast, the opposite is true for truthful examinees, and these differential concerns are manifested in the relative magnitude of the physiological reactions. The differences found between the magnitude of the physiological responses that accompany the relevant questions and those that appear in the comparison questions indicate the veracity of the examinee about the relevant questions but does not mean that the veracity of the subject's answers to the relevant questions is changing with any change of the detected difference. The subject's veracity stays the same because it does not depend on the test but on the subjective cognition of the examinee. What is changing is the degree one can rely on the detected difference to reflect the truthfulness of the examinee's answers to the relevant questions. Looking for essential consistency in that matter means repeating the measuring of the difference in reactions between R and C questions to determine whether the detected difference keeps pointing in the same direction.

In the field of test theories, consistency is closely related to reliability. If a measurement device or measurement procedure consistently assigns the same score to individuals or objects with equal values, they are considered reliable, but if the scores assigned to the same people or objects vary in repeated measurements in the absence of any known and understandable reason, the device or procedure is considered unreliable. Reliability is the extent to which a test or any measuring procedure yields the same results on repeated measures or trials or the extent to which a repeated test yields consistent outcomes and scores. Thus, reliability can be defined by the consistency of a measurement procedure and its outcomes. Reliability is the extent to which a test or any measuring procedure yields the same results on repeated measures or trials or the proportion to which a repeated test yields consistent outcomes and scores. Thus, reliability can be defined by the consistency of a measurement procedure and its outcomes. In CQT (and other polygraph techniques), consistency is not just a desirable option to look for but a critical factor for decision-making. We want to ensure that the outcomes we get on the test using our measuring procedure represent the examinees' actual mental states regarding their relative concerns about the relevant versus the comparison questions and not mere reflections of random fluctuations or other nonstable irrelevant factors.

The desire to differentiate the true state of a person who takes the test in the variable we are interested in measuring, from non-relevant influences that do exist in any test can be found in the "True Score Theory" (Lord, Novick & Birnbaum, 1968; Trochim, 2000) that, in its simplified version, when adapted to our field, states that – the measured physiological responses to any question and the measured difference between the strength of the responses to relevant (R) vs. comparison (C) questions, are a combination of "True score" plus "Error." That is to say that, in principle, the output itself may always contain irrelevant aspects, and the measurement device or the act of measuring might add another irrelevant value, namely Measurement Error. Thus the polygraph test variance includes true variability in the valence held by the examinees plus variance that reflects the effect of irrelevant factors[6]. For example, a person's test score might deviate from the true score because they are sick, anxious, in a noisy room, inattentive, got stomach cramps, are distracted by the examiner's inappropriate behavior, etc.

A significant portion of these irrelevant factors is the existence of random fluctuations, for instance, in the momentary level of the examinee's attention capacity or focus and in the pattern and the strength of the psycho-physiological reactions. These fluctuations might mask the true values and interfere with making the right decision. The need to nullify or at least to weaken the effect of these fluctuations is crucial for any test, and a polygraph test is no exception. Repeating the questions and looking for consistency over the repetitions is the primary way to achieve that.

The logic behind this tactic is that random fluctuations tend to nullify themselves in repeated measures and enable the true score to surface. The degree to which a particular technique or test is immune to the effects of such fluctuations on the outcomes is defined by the Reliability index that takes values between 0.0 to 1.0, indicating the extent to which an assessment tool or procedure is consistent, i.e., free from random measurement error. The closer the computed index is to 1.0, the higher the test output's consistency and the lower the noise's final effect.

To nullify or reduce the weight of the random errors factor that masks the true score takes repeating the measurements, and as long as its effects are presumed to exist, the Inconclusive Zone assumes the role of being the primary buffer from its adverse effects.

---

[6] It is important to understand that the variable here is not deception vs. truth-telling rather it is the difference between the reactions to the relevant and the comparison questions that supposed to reflect the difference in concerns between the questions. This difference can take various values which are the combination of true score (i.e. reflecting the amount of concern) plus error. There are two kinds of errors, random and systematic. Striving for consistency deals with the effort to reduce random error but can not remedy the effect of possible systematic error.

Having stressed the crucial role of repetitions, one may wonder about the number of repetitions needed to gain confidence that the outcome is reliable. Most techniques would ask for three repeated tests or three charts. Is that enough? Given the fact that in numerical scoring, as for now, it is customary to give the same weight to whichever repetition and every comparison point[7], using two charts with three relevant questions equals three charts with two relevant questions (in a single issue, single facet examination). In light of this, one may wonder whether this equality represents a latent unintended and surely undeclared attitude toward question repetitions which is rooted deeply in qualitative considerations far beyond the numerical scoring technique.

The role of having multiple comparison points

Multiple comparison points help expose relevant factor values and eliminate non-relevant ones, such as random fluctuations, or possible systematic irrelevant noise effects, such as the order by which the questions are presented.

By repeating the measuring of responses to the repeated stimuli, the random fluctuations of the noise tend to cancel each other and change the order of questions between repetitions to nullify possible systematic irrelevant noise. In contrast, the pertinent signals, i.e., the values of the responses to the "pure" stimulus that usually points to the same direction, surfaced above the noise. That is the theoretical reason for repeating the questions several times. Whenever the number of measurements is small, the relative effect of the noise on the variance or, in other words, the weight of the noise effect in the measured activity relative to the pertinent signal tends to be pretty strong, so there is a need to look for robust signals to be sure that what is detected is a pertinent signal and not a random fluctuation of the noise. However, when applying more measuring points, the noise effect is reduced, and even a relatively weak signal can surface above the noise and be detected. This should affect the size of the Inconclusive Zone. The smaller the number of comparison points, the bigger the minimum signal value to which significance can be attached, which means an increased Inconclusive Zone. Whereas having numerous comparison points, one can be satisfied with a much narrower Inconclusive Zone since after nullifying most of the random noise, even a moderate signal is enough to be consequential. This seems to be at odds with what some polygraph experts think, as they indirectly claimed that when using fewer comparison points (e.g., two relevant

[7] Some scoring methods, e.g., ESS or OSS3, double the EDA weight.

questions instead of three relevant ones), a narrower Inconclusive Zone is enough to distinguish DI from NDI calls because the maximum possible score is potentially smaller than the maximum possible score when more comparison points are used. That has falsely led them to opposite conclusions concerning the required size of the Inconclusive Zone for preventing errors due to random or non-random noise[8].

Exact repetitions of stimuli are the best technique to eliminate random noise since the noise tends to nullify itself due to the random quality of its changing directions and intensities. Ideally, in the end, we are left with the non-random effects that tend to pile up each time it is measured. However, there are also non-random irrelevant stimuli and responses that grow to build up too with repetitions and can add or subtract their values to the pertinent signals depending on their direction. There are several kinds of non-random irrelevant stimuli that we would like to reduce their impact. One type encompasses some characteristics embedded in the relevant questions that prompt concern in the truthful examinees from reasons other than being deceptive, such as being ashamed or embarrassed by the way the questions have been phrased or concerned about possible public exposure of the questions. Another type is emotionally loaded characteristics that do not induce concern but do activate the physiological system pretty much the same, such as mentioning in the questions specific names or circumstances. A third type is known as an outside-issue matter which mostly means that something outside the scope of the test or the Relevant and Comparison questions spheres bothers the examinees, not allowing them to direct enough attention to the R or C questions. Then, of course, there is always the general non-specific tension induced by the testing situation and its procedure and the concern of being diagnosed mistakenly. This does not exhaust all possibilities of directly affecting the examinees, but for now, I would like to enlighten another angle of possible effects of irrelevant noises. That type of noise affects the examiners, and only through them, occasionally, might roll onto the examinees' responses. An example of that can be the contamination by prior information about the guilt or innocence of the examinee.

The best way to lower the adverse effects of such non-random noise is by building a controlled testing atmosphere and procedure that prevents them from showing up. And in case they do show up, to speed the pace of their habituation with repetitions. Unfortunately, speeding up habituations of non-random irrelevant contamination without affecting the relevant stimuli is a matter that needs some sophisti-

[8] Personal written communication with anonymous reviewers and the chief editor of the Polygraph & Forensic Credibility Assessment: A Journal of Science and Field Practice, 2022.

cation that is beyond the material learned in polygraph schools and may contradict the policy of moving towards stiff standardization across cases, circumstances, and individual differences between examinees.

Unfortunately, even applying all these precautions is never enough to ensure that the effects of the irrelevant factors have been nullified. We should assume that the final measure contains a bit of random noise and, more dangerously, some non-random interference. Thus, the reason for applying an Inconclusive Zone is twofold. First, to buffer the random fluctuation remains and second, to function as a safeguard against the adverse effect of non-random variables hoping that we have managed to weaken their impact to such a low level that they would not cross the Inconclusive Zone barrier. Of course, there is nothing sacred in any size or shape of an Inconclusive Zone, but we should follow certain logic and steps in opting for the optimal zone.

The pay-off matrix of the inconclusive rate vs. errors and the optimal Inconclusive Zone

The general role of the Inconclusive Zone is to prevent wrong decisions. However, it comes with an inherent cost tag of reducing the number of correct decisions. Almost by definition, there is a negative correlation between the size of the Inconclusive Zone and the number of definite calls made by the test. The number of both kinds of decisions, whether wrong or correct, decreased whenever the range of the Inconclusive Zone increased. So, a wise approach is to think through a pay-off matrix and look for the optimal zone breadth. However, the optimum is not a fixed value; what might be regarded as optimal in one circumstance might not suit a different situation. Sticking to a rigid size of an Inconclusive Zone is like saying that all decisions have similar importance and consequences. Ideally, it should be left to the clients – who use the polygraph as a means to aid their decision-making to decide directly or undirectly how big the Inconclusive Zone should be. The case of homeland security is quite different from catching a thief, and so is the difference between sexual harassment and cheating on a school examination. This is a matter of social and institutional preferences based on economic, moral, and ethical considerations. Still, an evidence-based approach is essential to enable reasonable assessments of the proper range of the Inconclusive Zone. Alas, due to the lack of data, it is inconceivable to expect evidence to support such a differential approach with high resolution. That leaves us with a less-than-perfect choice to base the decision about the appropriate size of the Inconclusive Zone mostly on gross estimation developed from insufficient data and conventions stemming from reason and intuition rather than factual data.

For years under the influence of a variety of American government and law enforcement polygraph institutions, supported by the Utah group led by David Raskin, the Inconclusive Zone was symmetrical around Zero though there were indications that probably an asymmetrical zone would function better as was recommended by Backster (1963), (e.g. Elaad, 1999; Krapohl & McManus, 1999). As for now, some polygraph formats recommends symmetrical Zone, whereas other have adopted asymmetrical Zone based on research and some pay-off matrix considerations (Krapohl & Shaw, 2015).

Usually, the pay-off matrix, in its beneficial side, deals with improving the accuracy in polygraph testing, i.e., increasing the percent of correct decisions (True Positive and True Negative). However, polygraph testing also includes the idea of fairness that describes the wish to give any examinee a fair and equal chance to prove their innocence (True Negative) and, in the same vein, not to let any examinees to have a better chance than others to pass the test if they were lying (False Negative error).

By definition, random fluctuations do not take sides. They do not incline to one side. Thus, using an asymmetrical Inconclusive Zone can not be a remedy to cope with its potentially adverse effect on decisions. Yet, it was demonstrated that using an asymmetrical Inconclusive Zone in some circumstances may improve accuracy by reducing the rate of errors (Elaad, 1999; Krapohl, 2005). So, if it is not a matter of coping with random fluctuations, it is probably a way to compensate for some weaknesses or deficiencies in the neat rationale underlying the CQT and/or its derived rules resulting in asymmetrical pressure to augment responses to one type of questions. For example, suppose it is found that, as a general tendency in some circumstances, the reactions to R questions are more robust than the comparison ones. This may violate the fairness principle against innocent subjects and increase the False Positive rate. In such a case, an asymmetrical Inconclusive Zone is meant to correct for inherent basic inclination towards one side and evenly split the expected error rates between false-positive and false-negative types. Thus, keeping the fairness principle by applying an asymmetrical Inconclusive Zone usually increases accuracy. However, this is not necessarily the result of applying an asymmetrical inconclusive zone, and sometimes it may even reduce the overall accuracy. Yet, one can adopt it as a policy if he cares to preserve the fairness principle and choose the balance between accuracy and fairness in using the CQT accordingly.

Another way to describe the importance of applying an asymmetrical Inconclusive Zone is to remember that the polygraph is not a means of discovering the truth for its own sake; instead, it is a means to improve decision-making whenever a decision is needed, and as such, the asymmetrical Inconclusive Zone can also serve policy that weighs the cost of one type of error higher than the other. For instance, the police would not like to miss a serial rapist to a greater degree than it cares about casting false suspicion on someone who, by further investigation, can prove his innocence.

Conclusions

The Inconclusive Zone is not part of Polygraph theory in general and the Comparison Question Test in particular. This assertion is valid no matter which theoretical framework of CQT is adopted. Be it the basic FFF notion with its distinct Fear branches – fear of lie detection and fear of possible consequences – that Reids used to embrace (1977) and are still used in the polygraph theoretical sphere for almost a hundred years or so[9]; The Backster's Psychological Set (1963); Ginton's Relevant Issue Gravity (RIG) (2009); Senter et al.'s Differential Salience (2010); Palmatier & Rovner's Preliminary Process Theory (2015); (all five are mentioned in Krapohl & Shaw, 2015: 201–204); Hont's Cognitive Load (2014), or any other theoretical framework.

The Inconclusive Zone does not explain anything about how it is possible to conclude from physiological measures about the state of mind or the awareness of an examinee regarding the truthfulness of their accounts. So, it is possible to ignore it and still be within the sphere that enables inference of truth and lies by measuring physiological activity. However, the Inconclusive Zone is an important practical concept in applying the test since, in any measuring situation, there is a built-in component of several kinds of errors, as has been suggested in the "True Score Theory" that asserts that any test score or measurement consisted of the "True Score" and "Error(s)" (Lord, Novick & Birnbaum, 1968; Trochim, 2000). The Inconclusive Zone aims to reduce the impact of potential errors, mainly derived from random noise and the errors derived from irrelevant systematic noise, that is left due to failure to nullify such factor by applying specific procedures and means. This is not the only means for reducing potential errors but by no means

[9] The author is not awared of any reliable document that attributes the first use of this notion in polygraph testing to any scientist or practioner.

is a primary one. It should be recognized that there is a negative relationship between the size of the Inconclusive Zone and the number of decisions (right or wrong) that can b taken. Thus, the right way to decide about the Zone's size and shape (symmetrical or asymmetrical) should follow thinking through a pay-off matrix. Finally, if one takes the polygraph as a decision-making aid and not as a "truth machine" for the sake of getting the truth per-se, it is wise to consider the non-decisive area with flexibility that follows the importance of the specific decision to be made.

References

Backster C. (1962), Methods of strengthening our polygraph technique. *Police*, 6, 61–68.

Backster C. (1963), Polygraph professionalization through technique standardization. *Law and Order*, 11, 63–64.

Elaad E. (1985), Decision Rules in Polygraph Examination. In: *Anti-terrorism, Forensic Science, Psychology in Police Investigations* (pp. 167–179). A Book of proceeding. First Published, 1985; Imprint Routledge. 2019. https://doi.org/10.4324/9780429036590.

Elaad E. (1999), The Control Question Technique: A search for improved decision rules. *Polygraph*, 28, 65–73.

Ginton A. (2009), Relevant Issue Gravity (RIG) strength – a new concept in PDD that reframes the notion of psychological set and the role of attention in CQT polygraph examinations. *Polygraph*, 38 (3), 204–201.

Ginton A. (2012), A non-standard method for estimating the accuracy of lie detection techniques demonstrated on a self-validating set of field polygraph examinations. *Psychology, Crime & Law*, 19, 577–594. https://doi.org/10.1080/1068316X.2013.765137

Ginton A. (2013), The Importance of the Consistency Factor in CQT and Other Polygraph Tests. *Polygraph*. 2013, 42, 146–162.

Honts C.R. (2014), Countermeasures and credibility assessment. In: D.C. Raskin, C.R. Honts & J.C. Kircher (eds), *Credibility assessment: Scientific research and applications* (pp. 131–158). Academic Press.

Krapohl D.J. (2005), Polygraph decision rules for evidentiary and paired-testing (Marin Protocol) applications. *Polygraph*, 34, 184–192.

Krapohl D.J. & McManus B. (1999), An objective method for manually scoring polygraph data. *Polygraph*, 29, 209–222.

Krapohl D.J. & Shaw P.K. (2015), *Fundamentals of Polygraph Practice*, Academic Press.

Lord F.M., Novick M.R. & Birnbaum A. (1968), *Statistical theories of mental Test scores*. Addison-Wesely, Oxford, England.

Matte J.A. (1996), *Forensic Psychophysiology using the polygraph*. JAM Publications, Williamsville N.Y.

Palmatier J.J. & Rovner L. (2015), Credibility assessment: Preliminary Process Theory, the polygraph process and construct validity. International, *Journal of Psychophysiology*, 95 (1), 3–13.

Raskin D.C. & Honts C.R. (2002), The comparison question test. In: M. Kleiner (ed.), *Handbook of Polygraph Testing* (pp. 1–48). Academic Press, New York.

Reid J.E.& Inbau F.B. (1977), *Truth and deception: The Polygraph ("Lie Detector") Technique.* Williams and Wilkins, Baltimore, MD.

Senter S.M., Weatherman D., Krapohl D.J. & Horvath F.S. (2010), Psychological set or differential salience: A proposal for reconciling theory and terminology in polygraph testing. *Polygraph*, 39(2), 109–117.

Trochim W.M. (2000), *The Research Methods Knowledge Base*, 2nd edition. Atomic Dog Publishing, Cincinnati, OH.

Vrij A. (2008), Detecting lies and deceit: Pitfalls and opportunities, 2nd edition. John Wiley & Sons Ltd. New York.

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

DOI: 10.2478/EP-2022-0009

The Effect of Training on the Effectiveness of Deception Detection

Jagoda Dzida
Adam Mickiewicz University Faculty of Law and Administration
jagoda.dzida@amu.edu.pl

Abstract

The research conducted aims at determining the impact of short training on the trainee's ability to detect deception with the use of non-instrumental methods of detection of deception. As the criminal procedure permits such methods of detection of deception, the results of the research are practical. The research was carried out on a group of participants given the task to watch a video recording and conclude whether the person presented told the truth or lied. Subsequently, the participants were given a short training on both verbal and non-verbal deception cues, illustrated with examples taken from the video they had watched. In the second leg of the test, the participants watched another video and decided whether the person presented was truthful or deceptive (on the grounds of deception cues they spotted). The results showed that the training improved participants' ability to detect deception by c. 22%.

Key words: deception detection, lie, training

1. Introduction

It is mostly desired ability to effectively distinguish the truth from lies. Moreover, such an ability is also important for the proper conduct of the criminal proceedings, due to the aim of the criminal procedure – to detect and punish the perpetrator. Determining, whether the witness' statements or the accused' explanations are true or lie, is inevitable to correctly identify the guilty ones and punish them for their acts.

During the criminal process it is also possible to use a polygraph, but with some restrictions (Grzegorczyk, 2014; Paprzycki, 2015) (e.g. it is not allowed to be used during the interrogation). However, there are no restrictions on non-instrumental methods of deception detection, like the observation of verbal and non-verbal behaviour. Therefore, it is important to improve the effectiveness of that method of deception detection due to its applicability – it cannot be done without the proper training on verbal and non-verbal behaviour for those who will be responsible for the assessment of the credibility of statements and explanations given before the court.

For the time being, no research on detection deception training has been provided in the Polish literature. Moreover, the personal experience of the Author provides that no such training is available for some vocational groups, like judges and lawyers, who are interested in deception detection during the criminal procedure.

Therefore, it is necessary to conduct a study in this area.

2. Cues of deception

Contrary to the most popular view, that there are no objective cues to deception in our behaviour (Vrij, Mann, 2001; Vrij, 2008), there are still some specific behaviours suggesting that someone is not telling the truth.

Such behaviours include micro-expressions, emblematic slips and improperly timed or asymmetric mimic expressions (Vrij, Mann, 2001: 187–203; Vrij, 2008; Ekman, Hager, Friesen, 1981; Haggard, Isaacs, 1966). Sadly, they are not always present during lying. The fact is, when they do appear in someone's behaviour, they do indicate deception. Therefore, it would be a mistake to try to detect deception, based only on these cues to deception – as they are not always present during lies. Hopefully, they are not the only known cues to deception; there are also other symptoms of lies, but to successfully used them to detect deception, we always have to take into consideration the situational context and personal habits or manners of the potential liar (Ekman, 2013; Widacki, Mirska, Wrońska, 2012).

This is why also so-called subjective cues of deception (Ulatowska, 2009) should be used during our assessment – not only do they increase the effectiveness of deception detection, but are also involuntary for the liar. It means that widely known cues of deception (like avoiding the stare of the interlocutor) do not have to be wrong – to use them, we simply need to include the normal behaviour of the person who is possibly lying and the situational context. For example, it is true that during lying we can observe a decrease in illustrators (Ekman, 1981: 269–278; Ekman, 2013: 93–99). However, the said dependence is actual for most but not for all liars – some people will start to illustrate more than they usually do. It is caused not only by inter-individual differences but also by the fact that liars are among us and they are perfectly aware of the cues of deception, so they can mask (Ekman, 2013: 298–310) their lies. Therefore, we should become suspicious not only when we spot any of the mentioned cues of deception but also when we observe the extraordinary 'genuine' behaviour (strictly avoiding any gesture that may suggest lying) – such behaviour will look artificial and over-controlled.

To sum it up, effective deception detection is an extremely difficult task. Moreover, some studies suggest that finding clues to deception is almost impossible (Brennen, Magnussen, 2020; Vrij, Hartwig, Granhag, 2019) and we can rely on none of the known signs of deception. Despite that, some studies proved that effective deception detection, based only on behavioural observation, is possible – with an effectiveness of about 80–100% (Ekman, 2013: 290–310; Vrij, 2009: 89–96; Vrij, Mann, 2001: 187–203). The training usually increased the effectiveness of deception detection by about 10%, depending on the training method (Vrij, 2009). The described studies were conducted either on professionals, vocationally dedicated to deception detection (e.g. policemen, prosecutors) or on laymen, unmotivated in detecting deception due to no vocational need in this area. It is a potential mistake, because it is vital to conduct a study on those, who will take up such activity in their future job (like, for example, future lawyers). There are also some controversies aroused around such training, as we do not know the exact procedure of the training, which was highlighted by Charles F. Bond (2008); some Authors do even suggest that such training on non-verbal cues of deception does not improve lie-detecting skills at all (Jordan et al., 2019).

Because of that, the experiment on the effect of training on the effectiveness of deception detection has been carried out, focusing on training law students (future lawyers, potential prosecutors, and judges), about both verbal and non-verbal cues to deception. The main hypothesis was that even a short training should improve the ability of participants to detect deception and that the increase should be at least as big as it was proved in the aforementioned literature (at least by 10%).

3. Material and Methods

Participants

The research was conducted on students of the Law and Administration Department of the Adam Mickiewicz University (Poznań), 5 men, and 8 women aged 20–24. All participants volunteered for the research.

In the preparations for the research, 3 more law students from the same Department took part in recordings (these students will be referred to as 'Actors'). The Actors were not known to the participants.

All students volunteered for the research. The small number of participants was limited by the availability of students highly interested in deception detection and motivated during training; less motivated students would not be able to learn deception detection during short-time training.

Materials

In the research, 2 different recordings (audio-video recordings) were used, depicting an Actor answering questions. An Actor was standing by the desk, with his face towards the camera – he and his body language were perfectly visible from the waist up.

In each recording, the Actors were asked to choose one out of three cardboard boxes. They were informed that in the box they may find a pencil, a pen or earphones. Contrary to what the Actors were told, during the first recording ('Recording 1') every box contained earphones. The Actors' task was to convince the person asking them questions that in the box they chose was a pen or a pencil (so, in fact, they had to lie). During the second recording ('Recording 2') all the boxes were empty (which was surprising as it was not what they, potentially, could have found). Before choosing the box, the Actors were informed about their task –

they are to convince the person asking questions that in the box is a pen, a pencil or earphones, up to their liking (so they were sure that they can either lie or tell the truth, describing the real object from the box). The Actors were promised a prize for the person, who will successfully convince the person asking questions and the other two Actors (who were also present in the room during the recording).

The Actors were only to do their task (convince the person asking questions that the specific item has been found in the box) but they could either lie or tell the truth when answering the questions – some of them were possible to be answered correctly even if the main task was to lie about what is inside the box.

The task of the interviewer was not to react in any way if any cue of deception could be spotted; the interviewer had to keep a straight face and reveal no emotion.

During each of the two recordings, the Actors were answering 7 questions, about the item inside the box and their own, personal experiences, as they were to describe a similar item that is owned by them in real life:

1) What was your first thought when you opened the box?
2) What is the use of the item inside the box/ what could you associate the item from the box with?
3) Can you please describe the item from the box?
4) What is its colour/shape/pattern, how is it in touch/ is it heavy?
5) Do you have a similar item? If so, please, describe it; when did you buy it/how often do you make use of it?
6) Is the item in the box yellow/black/white?
7) Can you please describe once again, very precisely, the item from the box?

In response to each question, the Actors could, to their liking, tell the truth or lie. In practice, no Actor chose to lie in all questions; it was not also possible, to tell the truth in response to all questions, so in most cases, it was 50:50 (lie/truth), which was useful in the main part of the study when the Recordings were used.

The empty box instead of the expected item was meant to balance the experience of lying from the first part of the recording. The Actors, feeling too comfortable, might have been more relaxed in the second part and, therefore, show no or not enough cues of deception. Their surprise resulting from having an empty box was necessary to arouse some anxiety during answering questions in Recording 2 (as it increased the chances of showing cues of deception).

As a result of this part of the research, 6 one-and-half minute recordings were created – 3 coming from 'Recording 1' and 3 coming from 'Recording 2'. The said recordings were then used during the main part of the research.

Procedure

The participant, while assessing the truthfulness of the Actors on the recordings, were completing the survey concerning the words and behaviours of the Actors. Their task was not only to decide, whether each Actor told the truth or lied in response to the question (7 of them) but also to justify their choice by describing the specific behaviour of the Actor, leading to such an assessment (why was it true or false).

To assess the potential differences between 'Recording 1' and 'Recording 2', a similar survey was completed by the Actors, who were to decide whether their colleagues' behaviour in response to the questions was true or false (and why). As the Actors were not trained on the cues to deception, it was useful to find out, whether their effectiveness of deception detection during Recordings 1 and 2 will be the same or not (control group, untrained one). Moreover, the Actors were promised a small reward (university gadget) for both the most skilful liar and the most skilful lie detection, which kept them motivated during the recordings.

In the first part of the research, the participants were watching Recording 1, during which they were to assess the truthfulness of the Actors' responses to the questions and to complete the survey.

In the second part, they were given training on verbal and non-verbal cues of deception, both objective ones (like, for example, micro-expressions) and subjective ones (like avoiding eye contact).

During the training, the participants were told the most important cue of deception would be a change in someone's behaviour but that they should be also aware of the possibility of making a mistake in their judgment (and were also explained, why they can be mistaken). Moreover, because some objective cues of deception, like the aforementioned micro-expressions, could have been skipped or omitted while watching the recording, the participants were also told to focus on other cues (like asymmetry in mimical expressions – for example, a smile – that may be a cue of deception). The participants were guided about minding the context, voice tone, number of pauses, logic and consistency of speech and body language.

All of the cues that could have been spotted on Recording 1, were demonstrated to the participants along with the information if the answer to the question was true or false. These cues of deception were found by the Author when meticulously watching the Recordings – as the Author was the interviewer and knew when the Actor was lying, it was easy to spot them on the video and present them to the participants (in slow motion video for micro-expressions; as a photo when false smile; as a recording for a verbal cue to deception). The training took about 30 minutes, with 7 minutes devoted solely to the presentation of the cues to the deception present in Recording 1.

The last part of the research consisted of presenting Recording 2 to the participants, along with the completion of the survey (just like in the first part of the research).

The effectiveness of deception detection was measured by scoring the right answers of the participants in the survey. For each correct answer (that is, marking each false Actor's answer as a false one and each true answer as a true one, with an explanation of the cue of deception used for such an answer) one point was given to the participant. In the survey concerning Recording 1, it was possible to score 21 points (7 answers from the 3 Actors). The deception detection ability is connected not only with spotting the lie but also with differentiating the truth from it – this is why the correctness of all assessments, both marking the answer as true and as a lie, was taken into account.

To statistically measure the effectiveness of deception detection, Wilcoxon's paired data test with a significance level of $\alpha = 0{,}05$ from Statistica 12 (Statsoft) software has been used. Moreover, to test the differentiation of the validity of the assessments, the coefficient of variation was also calculated.

During Recording 2 there were similar cues of deception shown by the Actors, so it was possible to do the task similarly well as in Recording 1.

4. Results

The results of the survey completed by the Actors (control group, not trained on cues of deception) did not prove any relevant changes in deception detection – it was 62% for the survey completed during Recording 1, and 64% for the second survey.

The participants' effectiveness in detecting deception was 40% in Recording 1, and 62% in Recording 2. The effectiveness in Recording 2 was statistically significant ($z = 2{,}59$; $p = 0{,}0096$; Fig. 1). The training was the only factor significantly distinguishing the two surveys (as there was no difference in the situation presented in both Recordings – what was proved by the survey completed by the Actors). Therefore, the increase in the effectiveness of detecting deception was caused by the training, conducted after Recording 1.

The result of the survey taken after each Recording showed gross interpersonal differences in detecting deception between the participants, with a standard coefficient of variation (Fig. 1, Tab. 1).

Figure 1. The number of correct assessments given by the participants in both recordings

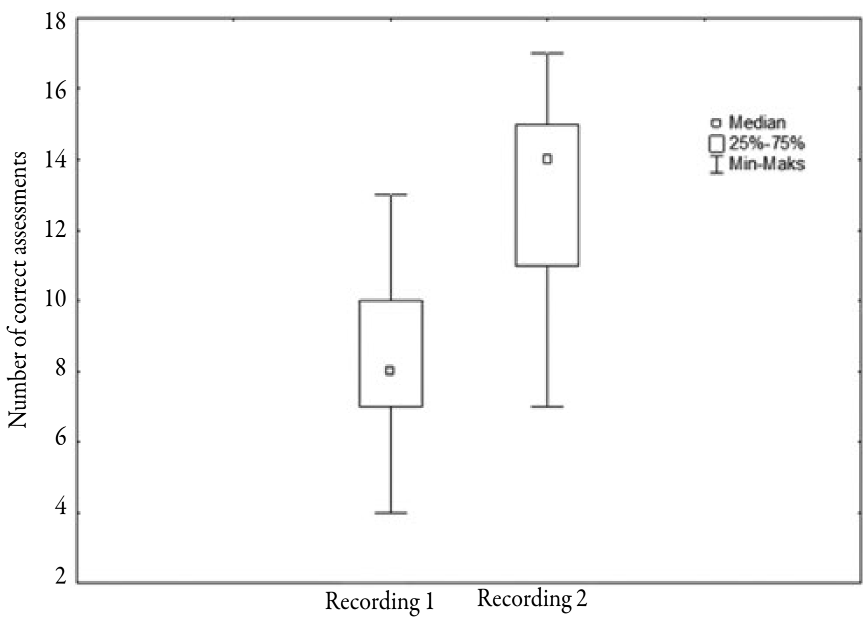

Table 1. The individual effectiveness of detecting deception of each participant

Participant	Number of correct assessments		Improvement (%)	Effectiveness of deception detection	
	Recording 1	Recording 2		Recording 1	Recording 2
1	8	14	75%	38%	67%
2	7	17	143%	33%	81%
3	10	8	-20%	48%	38%
4	11	9	-18%	52%	43%
5	11	15	36%	52%	71%
6	6	7	17%	29%	33%
7	7	15	114%	33%	71%
8	5	14	180%	24%	67%
9	7	13	86%	33%	62%
10	13	13	0%	62%	62%
11	9	11	22%	43%	52%
12	4	17	325%	19%	81%
13	10	16	60%	48%	76%
Average	8,3	13	78%	40%	62%
Range	4–13	7–17	-20% – 325%	19–62%	33–81%
Coefficient of variation	32%	26%	122%	32%	26%

5. The discussion

The study aimed to find out if the training on both verbal and non-verbal cues to deception may improve the effectiveness of detecting deception. As expected, in most cases such training may be beneficial to the effectiveness of deception detection.

In the future, it will be vital to conduct a similar study on a wider group of participants, using many recordings of more Actors than in the present study. The current results indicate that training on cues of deception should be provided – the effectiveness of detecting deception increased by about 22% thanks to such training, which verified the main hypothesis of the research. Moreover, the other researchers (de Turck, 1991: 81–89; de Turck et al., 1990: 189–199; Vrij, Graham, 1997: 144–148) on the subject also proved the increase in the effectiveness of deception detection due to training – by about 9–15%, with the initial effectiveness of detecting deception of about 42–55% (which is more than in the current research, as the students initially detected deception with the effectiveness of about 40%). The training of the professionals (policemen) results in a decrease in the effectiveness of deception detection (Köhnken, 1897; Vrij, 1994; Vrij, Graham 1997) – therefore it is vital to provide training for non-professionals (before they start to detect deception professionally in their work and gain their own experience in the area of lie detection); it would be beneficial perhaps to propose a new subject at law school, or train the apprentices before they start their working as a prosecutor, judge or lawyer.

It is impossible to refer to the other studies (Jordan et al., 2019), showing no difference between the trained and untrained groups – as it was previously stated (Bond, 2008), we have no control over the procedure of training that was provided, so it is possible that the training in these cases was improper, or that participants were unmotivated; we can only guess.

It must be clearly stated that in the research during the training, the participants were given the chance to get to know the behavioural cues to the deception of the Actors – they were given examples of such behaviour caught on Recording 1, after completing the survey as a part of the training. It might be a disadvantage of the research, as the participants would probably be less effective without getting to know better the behaviour of the person who is going to be assessed.

Yet, it does not make the research faulty or less valuable, as in the criminal process such technique is widely used – before the main part of the interrogation (the questioning), there is time for getting to know the person who is going to be questioned – in that part of interrogation an informal conversation can be made. Similarly, the interrogator during questioning asks questions to which he already knows the answers – it allows not only to control the truthfulness of the interrogated person but also to get to know his behaviour while lying or telling the truth. The participants were, therefore, put in a similar situation – they were shown the behaviour of the Actors while they were lying or telling the truth.

However, the huge differences in the individual effectiveness of detecting deception between participants are intriguing. One of the reasons may be the interpersonal differences – not everyone is predisposed to detect deception (like, for example, not everyone is perceptive enough). Moreover, not everyone may be studious and smart enough to be able to immediately and properly use the information given to him during training. It is also impossible to predict if each participant is equally eager to learn and to participate in a study, even if they volunteered for it; the participants might be tired, stressed or get uninterested during the research – it simply cannot be controlled. The additional aspect is their personal feelings towards the research, the training, the Actors or even the researcher – it may also affect their results of the effectiveness of detecting deception.

As for the initial effectiveness of the participants (before training) and the final one (after training), it is satisfactory – the results of the training are even better than expected (taking into account the aforementioned studies on the subject (with the increase in effectiveness exceeding 15%). Yet, it must be reminded that the said increase in effectiveness was possible because the training included the cues of deception previously recorded in Actors' behaviour. Possibly, if the participants were to assess the behaviour of other people (with whose behaviour they were not familiar), the increase would not be that significant.

The experiment proved that further research must be conducted in this area, including testing the effectiveness of deception detection with a lapse of time (long after the training). It should be verified, if the ability to detect deception remains on the same level or if the participants may forget about how to detect lies. In other words – whether such training causes some long-lasting effects or should be repeated after some time. Moreover, the research should be conducted on a larger group of participants and using a large number of Recordings (with many Actors). It could allow participants to better use their deception-detecting knowledge in practice and, at the same time, validate the results of the current study.

The results of the study are promising – the research proved that training non-professionals in deception detection may give much better and more satisfactory results than similar training, conducted on professionals. It also gives hope for further usage of non-instrumental methods of deception detection in the criminal process, which could potentially improve the detection of crimes as well as help to impose more adequate penalties and penal measures.

6. The limitations of the research

The aforementioned studies do have some limitations, which have already been mentioned before. The very first of them is a small research sample. Despite that, the participants were average students of the Faculty (the students of different years); each student has been watching several recordings; the results of the training were statistically significant.

The other limitation is using Recording 2, with the same Actors as previously seen in Recording 1. It is, then, possible, that the participants effectively detected deception only thanks to getting to know the Actors' behaviour in Recording 1. This is why further research in the area is needed, with the use of many Recordings with different Actors. However, during criminal interrogation in Poland, the policemen do have the ability to control the truthfulness of the interrogated person (as was already mentioned), so such a limitation of the research is not depreciating. Moreover, the display of the cues to deception shown by the Actors lasted for only 7 out of 30 minutes of training. Therefore, it is impossible to assume that the bigger part of the training was worthless and did not support the increase in the effectiveness of detecting deception.

References

Bond C.F. (2008), Commentary. A Few Can Catch a Liar, Sometimes: Comments on Ekman and O'Sullivan (1991), as well as Ekman, O'Sullivan, and Frank (1999). *Applied Cognitive Psychology*, December, https://doi.org/10.1002/acp.1475.

Brennen T., Magnussen S. (2020), Research on Non-verbal Signs of Lies and Deceit: A Blind Alley. *Frontiers in Psychology*, 14 December, https://doi.org/10.3389/fpsyg.2020.613410.

deTurck M.A. (1991), Training observers to detect spontaneous deception: effects of gender. *Communication Reports*, 4, 81–89.

deTurck M.A., Harszlak J.J., Bodhorn D.J. & Texter L.A. (1990), The effects of training social perceivers to detect deception from behavioral cues. *Communication Quarterly*, 38, 189–199.

Ekman P. (1981), Mistakes When Deceiving. *Annals of the New York Academy of Sciences*, 364, p. 269–278.

Ekman P. (2013), *Kłamstwo i jego wykrywanie w biznesie, polityce i małżeństwie*. Wydawnictwo Naukowe PWN, Warszawa, 146–170.

Ekman P., Hager J.C. & Friesen W.V. (1981), The Symmetry of Emotional and Deliberate Facial Actions. *Psychophysiology*, 18/2, 101–106.

Grzegorczyk T.H. (2014), Komentarz do art. 199(a). Zgoda osoby badanej na przeprowadzenie badań poligraficznych. In: *Kodeks postępowania karnego*. t. 1: *Artykuły 1-467. Komentarz*, LEX.

Haggard E.A. & Isaacs K.S. (1966), Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In: *Methods of Research in Psychotherapy*, 154–165.

Jordan S., Brimbal L., Wallace D.B., Kassin S.M., Hartwig M. & Street C.N.H. (2019), A test of the micro-expression training tool: does it improve lie detection? *Journal of Investigative Psychology and Offender Profiling* 16, 222–235, https://doi.org/10.1002/jip.1532.

Köhnken G. (1987), Training police officers to detect deceptive eyewitness statements. Does it work? *Social Behaviour*, 2, 1–17.

Paprzycki L.K. (red.) (2015), Komentarz do art. 199(a). In: *Komentarz aktualizowany do art. 1-424 Kodeksu postępowania karnego*. LEX.

Ulatowska J. (2009), Wiedza o wskaźnikach kłamstwa – w poszukiwaniu czynników, które na nią wpływają. *Problems of Forensic Sciences*, LXXX, 411–428.

Vrij A. (1994), The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior*, 18, 117–137.

Vrij A. (2008), *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley and Sons, Chichester 2008.

Vrij A. (2009), *Wykrywanie kłamstw i oszukiwania. Psychologia kłamania i konsekwencje dla praktyki zawodowej*. Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, 89–96.

Vrij A. & Graham S. (1997), Individual differences between liars and the ability to detect lies, Expert Evidence. *The International Digest of Human Behaviour Science and Law*, 5, 144–148.

Vrij A., Hartwig M. & Granhag P.A. (2019), Reading lies: non-verbal communication and deception. *Annual Review of Psychology*, 70, 294–317, https://doi.org/10.1146/annurev-psych-010418-103135.

Vrij A. & Mann S. (2001), Telling and Detecting Lies in a High-stake Situation. The Case of a Convicted Murderer, *Applied Cognitive Psychology*, 15, 187–203.

Widacki J., Mirska N. & Wrońska M. (2012), Werbalne i niewerbalne symptomy kłamstwa w oczach policjantów i psychologów, *Przegląd Bezpieczeństwa Wewnętrznego*, 7 (4), 19–31.

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

DOI: 10.2478/EP-2022-0010

Attempt at Detection of Deception Based on Records of Physiological Reactions Remotely Captured with FaceReader Software. Part 1

Jan Widacki
Andrzej Frycz Modrzewski Krakow University
jan.widacki@gmail.com

Bartosz Wójcik
Andrzej Frycz Modrzewski Krakow University

Anna Szuba-Boroń
Andrzej Frycz Modrzewski Krakow University

Abstract

Strong emotions are among others manifested in the expressive movements (facial expression). Facial expressions are natural and universal by nature. They do not depend on ethnicity, culture, social status, age, etc. Nonetheless, humans are sometimes capable of controlling their facial expressions and hiding their emotions. Simulating emotions is a fundamental acting skill. However, controlling facial impressions takes time. The onset of such a control is delayed by anything from 0.25 to even 0.1 second – the period when the authentic facial expression, adequate to the emotion is demonstrated – and therefore remains imperceptible to an external observer. This short-lived facial expression observed in that short meantime is known as microexpression.

FaceReader, designed by Dutch company Noldus (established and directed by Professor Lucas Noldus), is a software package for automatic recognition and analysis of facial expression. As its diagnostic value for validity as well as reliability, that is the level of correct indications, remains unknown, we decided to determine it experimentally and have chosen to run an experiment comparing its diagnostic value with that of a traditional polygraph examination.

Key words: FaceReader, FaceReader (Noldus), detection of deception, facial expressions, microexpression, Noldus

1. Introduction

Deception, lie, guilty knowledge, concealed information (terms that are identical from the methodological and logical points of view (Widacki, 2011), and therefore can be used interchangeably) are known to be accompanied by emotional changes and the subject's intellectual effort of self-control during a polygraph examination (Widacki, Dukała, 2015). Altogether, the reactions to the stimuli provided in the test questions trigger emotional changes (changes of body activity). Those changes of activity are accompanied by physiological changes (reactions) that are registered by a polygraph machine, and later interpreted in the context of the stimuli (test questions) that caused the reactions. The results of the interpretation can provide circumstantial evidence in criminal procedures, which, as any evidence, requires being confronted with the remaining body of evidence (Widacki, 2014).

Currently the effort of researchers of instrumental detection of deception runs basically along two main tracks. The first is the attempt to move from the psychophysiological level of the detection of deception to the neuropsychological one, using EEG and fMRI (Widacki, 2007).

The other, to which this study and research belong, is based on seeking such physiological correlates of body activity changes at the psychophysiological level that could be observed and recorded remotely (without contact), without the need to attach sensors to the body of the subject, which is the case in traditional polygraph examinations. It is a step towards the development of efficient methods of instrumental detection of deception without the knowledge, and ergo also consent, of the subject.

Such an examination, if undertaken without the knowledge and concession of the subject, obviously raises questions of ethical and legal nature, which require separate investigation and are not considered in this article. We generally believe that no such examinations should be carried out, and results obtained through them should not be permitted as evidence for investigative purposes, and even more so as evidence in criminal cases. However, such procedures seem useful in prevention of terrorist acts, and also in the operation of intelligence and counter-intelligence.

2. Attempts at lie detection based on remotely recorded physiological changes

Designing an efficient method of detection of deception based on observation, recording, and analysing data on physiological reactions obtained remotely first of all requires finding such reactions whose diagnostic value would be comparable to the diagnostic value of reactions recorded by a classical polygraph machine, that is reactions in the form of changes in the course of the breathing functions, the operation of the blood system, and the skin galvanic (electrodermal) responses. Secondly, which is furthermore far more difficult, it calls for designing such technique of examination in which the subject will not be aware that their physiological reactions are observed and recorded while they answer the questions. This makes it evident that the tests and types of questions used in this case must be different than in classical polygraph examinations.

State-of-the-art knowledge and technology make it possible to record remotely at least several physiological reactions that accompany deception. These include voice changes, eyeball movements, changes in facial temperature, and changes in facial expression. Each of these reactions, as well as attempts at their diagnostic use for detection of deception, has been subject of plentiful literature.

Voice changes were one of the first remotely observable parameters to be tested. Even without the use of any mechanical devices, the voice, not unlike facial expressions, provides grounds for recognising fundamental emotions, especially if powerfully experienced, as well as certain moods. Voice reflects fury, anger, fear, impatience, shyness, self-assuredness, and other emotions.

Attempts at lie detection based on voice changes have been undertaken in the United States since the 1970s. Several devices recording such changes have made it to the market under various brand names. Initially they were the PSE (Psychological

Stress Evaluator) and the VSA (Voice Stress Analizer), and later the CVSA (Computerised Voice Stress Analyzer) (Widacki, 2018). A similar device under the name of LVA 6.50 was constructed in Israel (Gramatyka, 2008).

The results were, however, discouraging, as the diagnostic value of detection of deception performed in this method remained far below the diagnostic value offered by polygraph examinations (Barland, 2002; Horvath, 2002). For that reason, they were disqualified by the US Department of Defence and American Equal Opportunity Commission that forbade their use in pre-employment tests, and also by the Wisconsin Court of Appeals (State of Wisconsin v. Paul D. Hoppe, 2001) (Widacki, 2018). Similarly, subsequent Polish experimental studies proved the relatively low diagnostic value of the tests conducted with such devices, remaining below the diagnostic value of polygraph examinations (Gramatyka, 2008; Pietruszka, 2006; 2008; 2009; Leśniak, Leśniak, Gramatyka, 2011).

Eyeball movement and pupillary reflex – attempts have long been made to use those as indicators of emotional changes, also ones accompanying lies. Avoiding eye contact and averting the eyes have been treated since times immemorial as symptoms of deception, most probably on the grounds of experience drawn from life. By the way, observation of eyeball movement is also used currently in non-instrumental lie detection (Gordon, Fleisher, 2011). It was only much later, towards the end of the 19th century that special devices, known as oculographs, began to be constructed for the observation of eyeball movement. Eye tracking, consisting of observation, recording, and analysis of eyeball movements, is used to determine what the eye focuses on. It is used in medicine, psychology (especially psychology of advertising) marketing, pedagogy, ergonomics, and studies of human–computer interactions among the others. Other than the eyeball movement, oculographs can also observe and record changes in the size of the subject's pupil (Duchowski, 2004; Jacob, Karn, 2003; Stolecka-Makowska, Wolny, 2014; Paśko, 2017; Orquin, Holmqvist, 2018). Initially, oculographs were installed on the head of the subject. Currently, apart from such devices, also ones that observe eyeball and pupillary movements remotely are used. Early in the 21st-century, attempts were made to use such devices for studying emotions, and especially for the detection of deception (Hacker et al. 2014; Kircher, Raskin, 2016). The results of those attempts also seem encouraging, even if the level of correct lie detection remains below that of classical polygraph examinations (Kircher, Raskin, 2016), and the techniques of such experimental examinations have as yet not required installing any sensors on the body of the subject even if they made the subject stay close the oculograph, and their fully conscious participation in the examination, for which reason they do not make performing such test without the subject's knowledge possible.

Changes in facial temperature have long been considered a good indicator of emotional changes (Hilgard, 1967: 243; Woodworth, Schlosberg, 1966: 268), however it was only the use of an infrared camera for their observation and recording that made it possible to perform such tests remotely (Gołaszewski, Zając, Widacki, 2015). The results of experiments detecting deception based on the observation of facial temperature changes using an infrared camera might have been encouraging, yet inferior to the attempts at such a detection made in parallel with the use of the polygraph (Widacki, Widacki, Antos, 2016; Mikrut, Widacki, Widacki, 2018). Moreover, those attempts proved that method of lie detection to be highly complex technologically, which made it impractical (Mikrut, Widacki, Widacki, 2018). It required close proximity of the infrared camera lens to the face of the subject, and by that token did not practically allow performing such an examination the subject realising it was performed.

Changes in facial expression (including "microexpression") – strong emotions are among others manifested in the expressive movements (facial expressions) (Woodworth, Schlosberg, 1966: 168). Elementary life experience makes it possible to read at least the fundamental and strongest emotions, notably anger (fury), fear, joy, surprise and the like from facial expressions. Psychologists differentiate from seven to nine basic emotions that are expressed by the face (see below).

Facial expressions are natural and the universal by nature. They do not depend on ethnicity, culture, social status, age, etc.

Nonetheless, humans are at times capable of mastering their facial expressions and hiding their emotions, just like they know how to simulate them. Simulating emotions is a fundamental acting skill. However, controlling facial impressions takes time. The time required for it is very short, in the range of 0.25–0.1 second, comparable to winking an eye, and therefore imperceptible for an external observer. This is the time when the authentic facial expression, adequate to the emotion, is demonstrated. This short-living facial expression is known as microexpression. As has been mentioned, its time is generally too short to be noticed by an external observer. That is why its recognition and recording requires special devices that allow capturing it and then replaying at a much slower speed. American psychologists Silvan Tomkins (1911–91) and Paul Ekman (b. 1934) are considered the discoverers of microexpression. The latter wrote many works on microexpression and detection of deception based on analysing microexpression. He has also written several books that have been translated among others into Polish (Ekman, 2011; 2001; 2003; Ekman, Davidson, 2012; Ekman, 2009; Ekman, Friesen, 1974; Ekman, Hager, 1979; Ekaman, Friesen, Ancoli, 1980). Ekman claims that studying microexpression is a good method of detecting deception (Ekman, 2001). Moreover, it seems that remote observation and recording of microexpression poses no major technical problems even if conducted without the subject's knowledge, unlike analysing facial temperature changes and eyeball movements.

As has been mentioned, when applied separately, all the methods described above (voice analysis, facial temperature change analysis, and microexpression analysis) have proven lower diagnostic values in experimental studies than polygraph examinations recording in parallel at least the following three functions of the organism: breathing, blood system, and the skin galvanic response (electrodermal reaction). As yet, there have been no results of studies combining the three other remotely recordable physiological reactions for detecting deception.

Of the remote methods for detection of deception listed above, it is perhaps the analysis of facial expressions, micro expressions included, that has recently received most attention. Many works devoted to the subject have lately been published all around the world (see: Samuel at al., 2019; Shen *et al.* 2021; Curtis, 2021; Ioannou *et al.* 2005; Dimberg, Thunberg, Grunedal*,* 2002; Monaro *et al.* 2022). In Poland this issue is practically unknown, and has only been discussed in a handful of general works that, furthermore, as a rule ignore most foreign literature (see: Laszczak, 2021).

Studying microexpression for capturing emotions serves many purposes, and is used primarily in psychology and psychiatry, and especially in the psychology of advertising and marketing. Its use for the detection of deception is only one of the potential options.

3. The mechanism of microexpression

The mechanism of microexpression is complex. At the level of execution, it is performed by 24 muscles of the head, mostly by the muscles of the face (*musculi capitis*). Around 70% of these muscles have been discovered to have no practical use save for the actuating facial expressions, which are known to express emotions.

Various facial expressions correspond to various emotional states. Ekman and Friesen (Ekman, Friesen, 1978), following Hjortsjö (1970) differentiate seven fundamental emotions:

- happiness
- anger
- fear
- surprise
- sadness
- disdain, and
- disgust.

Other authors include other fundamental emotions such as shame or revulsion (Tomkins 1999; 2008), and interest and surprise (Izard, 1977; 1994; Izard, Rosen, 1998).

A different muscular complex corresponds to each of these emotions. For example, drawing the brows towards each other and downwards, increasing the distance between the eyelids, pressing the lips, tilting the head slightly forward, and optionally also pushing the jaw forward are characteristic of the emotion of anger. The emotion of anger was intended to get the body ready to fight. In turn, the emotion of fear was to prepare the body for flight (Cannon, 1932: 227; Kępiński, 1972). In these cases, facial expressions played several functions, both independent (e.g. to scare the opponent) and derived from other changes in the body, and adjusting the body to operation in new conditions.

4. Observation of microexpression

By the very nature of facial expression changes, their observation calls for paying attention to several changes that take place in the subject's face simultaneously. In case of microexpressions, their aforementioned extremely short duration, lasting from 0.25 to 0.1 second, provides an additional hurdle.

There have been numerous attempts at solving the problem, much like there have been many attempts at investigation of microexpression for the detection of deception. As far as the first goes, attempts have been undertaken to develop special computer software to automatically investigate in parallel multiple elements of facial expressions and transform them into a legible chart, making it possible to recognise precisely the emotion causing facial changes.

Therefore, it is little wonder that such works are primarily conducted in teams with participation of computer scientists and in such departments of, usually, universities of technology as computer engineering, electronics and mechatronics, and computer sciences (Ioannou, Caridacis, Karpousis, Kollias, 2006; Starostenko, Contreras *et al.* 2011).

5. FaceReader system from Noldus

One of such software packages is FaceReader designed by Dutch company Noldus (established and directed by Professor Lucas Noldus of Radboud University in the Netherlands) that allows remote observation of the subject's facial expressions, processing information obtained from that observation, and returning their averaged recording as it recognises emotions by analysing the facial expressions (expressing emotions).

Thus, FaceReader™ is a software package for automatic recognition and analysis of facial expressions, notably the expressions of the six fundamental (universal) emotions: happiness, sadness, anger, surprise, fear, and disgust.

The software makes use of FACS (Facial Action Coding System) created by a Swiss anatomist Carl Herman Hjortsjö (1970), developed by Paul Ekman and Wallace V. Friesen, and published in 1978 (Ekman, Friesen, 1978) to be updated further later (Ekman, Friesen, Hager, 2002). The system "orders" the movements of facial muscles and allows categorising the movements of facial muscles for more precise determination of the emotional state.

The main task of FaceReader is to classify the facial expressions of participants in the test. The results obtained are analysed and visualised in real-time on several charts. These include line and/or bar charts, or alternatively pie charts showing the percentage share of a specific emotion in the process of expression.

Determining the facial expression with FaceReader:

1. **Face finding** – the position of the face in an image is found using a deep-learning-based face-finding algorithm, which searches for areas in the image having the appearance of a face at different scales.

2. **Face modelling** – FaceReader uses a facial modelling technique based on deep neural networks. It synthesises an artificial face model, which describes the location

of 468 key points in the face. It is a single pass quick method to directly estimate the full collection of landmarks in the face. After the initial estimation, the key points are compressed using Principal Component Analysis. This leads to a highly compressed vector representation describing the state of the face.

3. **Face classification** – with the above provided, classification on the facial expressions takes place by a trained deep artificial neural network recognising patterns in the face. FaceReader directly classifies the facial expressions from image pixels. Over 20,000 manually annotated images were used to train the artificial neural network.

The model allows FaceReader to classify facial expressions, and record Action Units (AU) and levels of activation, while the valency of facial expressions can be visualised.

Additionally, FaceReader is capable of discerning the "neutral" state of emotions and "contempt". It also records:

- blinking frequency (separately for the left and right eye)
- Head positions: "head turn left" (AU 51), "head turn right" (AU 52), "head up" (AU 53), "head down" (AU 54), "head tilt left" (AU 55), and "head tilt right" (AU 56)
- direction of gaze
- other features including gender, age, and facial hair (beard and/or moustache)
- heart rate/heart rate variations, in individually defined complex expression standard.

Other independent variables can be additionally introduced manually.

It must be noted that FaceReader analyses the left and right Action Units (AUs) separately. This unique feature makes it possible to define the intensity of muscle activity separately for the left and right sides of the face.

FaceReader captures and analyses 20 Action Units as well as several frequent and complex combinations thereof:

AU 1. – inner brow raiser

AU 2. – outer brow raiser

AU 4. – brow lowerer

AU 5. – upper lid raiser

AU 6. – cheek raiser

AU 7. – lid tightener
AU 9. – nose wrinkler
AU 10. – upper lip raiser
AU 12. – lip corner puller
AU 14. – dimpler
AU 15. – lip corner depressor
AU 17. – chin raiser
AU 18. – lip pucker
AU 20. – lip stretcher
AU 23. – lip tightener
AU 24. – lip pressor
AU 25. – lips part
AU 26. – jaw drop
AU 27. – mouth stretch, and
AU 43. – eyes closed.

The preliminary studies in the possibility of using computer applications for analysing emotions in studying deception in psychotherapy have already been conducted (Curtis, 2021). The capacity technology offers in investigating and understanding behaviour is emphasised in the context of multiple scientific disciplines. FaceReader software has been tested preliminary in consumer research (efficiency of advertising) and also in studying emotions in children based on video recordings to compare lie detecting efficiency of FaceReader and adults talking to the child (Gadea *et al.* 2015).

As the software producer (Noldus) informed, save for the case quoted above, FaceReader has never been used for the "detection of deception".

The diagnostic value of FaceReader, that is the level of its correct indications, remains unknown and we want to find it experimentally. Our experiment will be conducted in multiple stages. The first will be to compare the efficiency (diagnostic value, including validity and reliability) of detecting deception with the use of a traditional polygraph (using traditional polygraph examination techniques, that is CQT and CIT (GKT)) and the FaceReader system. The two methods will be applied in parallel. Thus, in the first stage of the experiment, the subject of a polygraph examination will have their facial expressions additionally captured and examined by FaceReader. The reactions captured by the polygraph will subsequently be compared to those captured by the FaceReader system.

In the following stage of the experiment, we envisage the application of CQT, CIT/GKT polygraph examination techniques, while the diagnosis will only be based on observation of facial expressions with the FaceReader system from Noldus. The subject will be aware of being examined and of detection of the potential deception being the purpose of the examination.

These experiments should jointly provide an answer to the question whether Noldus's FaceReader can be successfully used in the place of a traditional polygraph machine.

If this is the case, that is the diagnostic value of such examination will be proven similar to that of a polygraph examination, we will proceed to the following, last stage of the experiment, which will have the examination remodelled so that the subject will not know that they are being examined and will not be informed about the examination. There will be no traditional CQ, CIT/GKT techniques used, but instead, a custom-built test with particular questions melded into a conversation, hidden between the non-diagnostic fragments of the conversation, will be tried out.

References

Barland G. (2002), Use of voice changes in the detection of deception. *Polygraph,* 31, 2.

Cannon W. (1932), *The wisdom of the body*. Northon & Comp. Inc.

Curtis D.A. (2021), Deception detection and emotion recognition: Investigating F.A.C.E software. *Psychotherapy Research*, 31, 6, 802–816.

Dimberg U., Thunberg M. & Grunedal S. (2002), Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition and Emotion,* 16 (4), 449–471.

Duchowski A.T. (2004), *Tracking Methodology: Theory and Practice*. Springer, London.

Ekman P. (2001), *Telling lies. Clues to deceit in marketplace, politics, and marriage*. W.W. Norton.

Ekman P. (2003), *Kłamstwo i jego wykrywanie w biznesie, polityce i małżeństwie*, tłum. polskie. PWN, Warszawa.

Ekman P. (2009), Lie Catching and Microexpressions. In: *The Philosophy of Deception*, ed. C.W. Martin. Oxford University Press, NY, 118–138.

Ekman P. & Davidson R.J. (eds) (2012), *Natura emocji. Podstawowe zagadnienia*. Gdańskie Wydawnictwo Psychologiczne, Sopot.

Ekman P. (2011), *Emocje ujawnione. Odkryj co ludzie chcą przed tobą zataić i dowiedz się czegoś więcej o sobie*, Wydawnictwo Sensus, Gliwice.

Ekman P. & Friesen W.V. (1974), Nonverbal behavior and psychopathology. In: *Psychology of depression: contemporary theory and research*, R.J. Friedman, M.N. Katz. Inston, Washington.

Ekman P. & Friesen W.V. (1978), *Facial action coding system: manual*. Consulting Psychologists Press, Palo Alto.

Ekman P., Friesen W.V. & Hager J.C. (2002), *Facial action coding system. The Manuel on CD-ROM. A human face*. University of Utah, Salt Lake City.

Ekman P. & Hager J.S. (1979), Long distance transmission of facial affect signals. *Ethology and Sociobiology*, 1, 78–82

Ekman P., Friesen W.V. & Ancoli S. (1980), Facial signs of emotional experience. *Journal of Personality and Social Psychology*, 39, 1125–1134.

Gadea M., Aliño M., Espert R. & Salvador A. (2015), Deceit and facial expression in children: the enabling role of the "poker face" child and the dependent personality of the detector. *Frontiers in Psychology*, 6.

Gołaszewski M., Zając P. & Widacki J. (2015), Thermal Vision as a Method of Detection of Deception. A Review of Experiences. *European Polygraph* 9, 1(31), 5–24.

Gordon N.J. & Fleisher W.L. (2011), *Effective Interviewing and Interrogation Techniques*, 3rd edition, Elselvier Science & Technology.

Gramatyka M. (2008), Teoretyczne i praktyczne aspekty analizy poziomu stresu w głosie ludzkim na przykładzie urządzenia LVA 6.50. In: *Kryminalistyka i nauki penalne wobec przestępczości. Księga pamiątkowa dedykowana Profesorowi Mirosławowi Owocowi*, ed. H. Kołecki. Wydawnictwo Poznańskie, Poznań, 367–377.

Hacker D.J., Kuhlman B., Kircher J.C., Cook A.E. & Woltz D.J. (2014), Detecting deception using ocular metrics during reading. In: *Credibility assessment: Scientific research and applications*, eds D.C. Raskin, C.R. Honts, J.C. Kircher. Elsevier.

Hilgard E.R. (1967), *Wprowadzenie do psychologii*. PWN, Warszawa.

Hjortsjö C.H. (1970), *Man's face and mimic language*. Nordens Boktryckeri, Malmö.

Horvath F. (2002), Experimental comparison of the psychological stress evaluator and the galvanic skin response in detection of deception. *Polygraph,* 31, 2.

Ioannou S.V., Raozaiou A. ZZT., Tzouuvaras V.A., Mailis T.P, Karapouzis K.C. & Kollias S.D. (2005), Emotion recognition through facial expression analysis based on a neurofuzzy network. *Neural Networks,* 18, 423–435.

Izard C.D. (1977), *Human emotions*. Plenum, New York.

Izard C.D. (1994), Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, 288–299.

Izard C.D. & Rosen J.B. (1998), Brain system, emotions and psychology. *Contemporary Psychology: A Journal of Reviews,* 43, 101–103.

Jacob R.J.K. & Karn K.S., (2003), Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. *Mind* 2 (8): 573–605.

Kępiński A. (1972), *Rytm życia*. Wydawnictwo Literackie, Kraków, 166–167.

Kircher J.C. & Raskin D.C. (2016), Laboratory and Field Research on the Ocular-motor Deception Test. *European Polygraph*, 10, 4 (38), 159–172.

Laszczak B. (2021), Ekspresja mimiczne – podstawowy katalog do wykrywania kłamstwa i emocji (część 1). *Prokuratura i Prawo*, 3, 144–161.

Leśniak M., Leśniak B. & Gramatyka M. (2011), Trafność wiarygodności wypowiedzi na podstawie wskazań analizatora głosu LVA 6.5. In: *Psychologiczne i interdyscyplinarne problemy w opiniodawstwie sądowym w sprawach cywilnych*, ed. J.M. Stanik. Katowice, 228–236.

Mikrut Z., Widacki M. & Widacki J. (2018), Experiments in Using Face Temperature Changes as an Indicator in Instrumental Detection of Deception. *European Polygraph,* 3, 85–105.

Mikrut Z., Widacki M. & Widacki J. (2018), Próba wykorzystania zmiany temperatury twarzy jako wskaźnika w instrumentalnej detekcji kłamstwa. In: *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa. Problemy kryminalistyczne, etyczne i prawne*, ed. J. Widacki, Kraków, 133–148.

Monaro M., Maldera S., Scarpazza C., Sarton G. & Navarin N., (2022), Detecting deception through facial expression in a dataset of videotaped interviews: A comparison between human judges and machine learning models, *Computers and Human Behavior*, 127, February 2022.

Orquin J.L. & Holmquist K. (2018), Threats to the validity of eye-movement research in Psychology. *Behavior Research Methods*, 50 (4), 1645–1656.

Paśko I. (2017), Eyetrackingowe badania we wczesnej edukacji przyrodniczej. *Pedagogika Przedszkolna i Wczesnoszkolna*, 5, 2/1, 205–215.

Pietruszka J. (2006), O perspektywach zastosowania analizatorów poziomu stresu w głosie. *Problemy Kryminalistyki,* 252, 64–72.

Pietruszka J. (2008), Wykorzystanie analizatora głosu (poligrafu – wariografu głosowego) w postępowaniu karnym i stosunkach pracy, unpublished doctoral dissertation, Katedra Kryminalistyki Uniwersytetu Warszawskiego, Warszawa.

Pietruszka J. (2009), A renaissance in voice analysers as tools for detection of deception? *European Polygraph,* 1(7), 33–37.

Samuel S.G., Chatterjee T., Thapliyal H. & Kacker P. (2019), Facial psychophysiology in forensic investigation: A novel idea for deception detection. *Journal of Forensic Dental Sciences,* 11, 2, 90–94.

Shen X., Fan G., Niu C. & Chen Zh. (2021), Catching a Liar Through Facial Expression of Fear. *Frontiers in Psychology,* 12, https://doi.org/10.3389/fpsyg.2021.675097.

Starostenko O., Contreras R., Aquina V.A., Pulido L.F., Asomoza J.R., Sergiyenko O. & Tyrsa V. (2011), A Fuzzy Reasoning Model for Recognition of Facial Expression, *Computación y Sistemas*, 15, 2, 163–180.

Stolecka A. & Wolny R. (2014), Możliwości zastosowania techniki okulograficznej w ilościowych badaniach marketingowych. *Zeszyty Naukowe UE w Katowicach. Studia Ekonomiczne* 195, 195–205.

Tomkins S.S. (1999), *Communicating emotions,* Cambridge University Press, Cambridge–London.

Tomkins S. S., (2008), Affect imagery consciousness, Springer Publ. Comp.

Widacki J. (2014), Results of polygraph examinations: direct or circumstantial evidence? *European Polygraph,* 8, 2(28): 61–67.

Widacki J. (2021), *History of polygraph examination*. Polskie Towarzystwo Kryminalistyczne.

Widacki J. (ed). (2018), *Kryminalistyka*, 4th ed. C.H. Beck, Warszawa, 457.

Widacki J. (2007), From forensic psychophysiology to forensic neurophysiology. New trends in examinations in the detection of deception. *European Polygraph,* 2.

Widacki J. (2011), Logical Identity of Conclusions from Polygraph Testing Performed in Control Questions Test (CQT) and Guilty Knowledge Test (GKT) Techniques. *European Polygraph,* 1(15), 6–10.

Widacki J. & Dukała K. (2015), Detekcja kłamstwa – czyli czego?, *Problemy kryminalistyki,* 287 (1), 3–16.

Widacki J. & Dukała K. (2015), Lie detection – of what?, *Problemy kryminalistyki*, 287 (1), 61–74.

Widacki J., Widacki M. & Antos J. (2016), Preparation to experimental testing of the potential from using facial temperature changes registered with an infrared camera in lie-detection. *European Polygraph,* 10, 1(35) 17–23.

Woodworth R.S. & Schlosberg H. (1966), *Psychologia eksperymentalna*. PWN, Warszawa, vol. 1.

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

DOI: 10.2478/EP-2022-0011

Discussion

In the previous issue, we published Jan Widacki's article entitled "What Do Polygraphers-Practitioners Expect from Science?" that, with encouragement from the editors, opened a discussion. Below, we publish Frank Horvat's voice in that discussion.

I am responding to the survey questions regarding the need for research on and in Polygraphy (see: „What Do Polygraphers–Practitioners Expect from Science?"). I will respond to each of your questions but I find it necessary to expand somewhat on this important topic. I've had a number of discussions on the topic in recent days with other examiners, including APA Board members, and find that my views seem to be somewhat divergent from theirs but in line with what seems to be your position. I'll try to keep my comments brief but will agree to write more in the event you find them of interest.

1. Do you believe that practical polygraphy needs cooperation with research and university centres?

YES Xo / NO o

I believe this is essential if we are to advance Polygraphy beyond its current state. What examiners alone are now doing is entirely inadequate. We need much more interest and innovation—and research—than is now evident.

2. Have currently conducted research, and its published results, been useful for your practice?

YES Xo / NO o But not just for my practice but for many others and, perhaps, for almost all practicing examiners and others with an interest in the field.

In response to question #3: I strongly favor more and better interaction between examiners, the APA and similar organizations, and university professors and researchers. As you know a number of years ago the APA provided support to my university, MSU, to arrange for the establishment of a Research Center devoted to research on topics related to Polygraphy. I directed that Center and used APA financial support to provide assistantships to graduate—and some undergraduate—students to carry out or assist in research projects. A number of projects were completed in the years the Center was active and most of them were published in either the APA publications or scientific journals. In my view the Center and the way it functioned was quite successful in spite of the fact that it had very limited financial support. But there was good support from principal persons at MSU and that was essential to its functioning. When I left the university for work at NCCA I was interested in maintaining the Center, at least the idea of it, by moving it to South Carolina. I had hoped to do this with a professor who had a strong background in fMRI and related technologies—and also an interest in Polygraphy. But, the APA decided not to proceed and withdrew from funding a Research Center. (I believe that there were some on the APA Board who were not interested in research and others who believed that what was already done was all that was necessary.) The APA funding, by the way, relative to typical research grants, was very, very small and not especially of great interest without strong support from one who would direct a Center. As a point of information my interest was in recruiting persons who were already trained and experienced examiners to do research. First, therefore, it was necessary to encourage such persons to apply for admission to a graduate program; that, in itself, was a challenge. This is not something that typically has been done at other universities where research in Polygraphy has been carried out.

I have served on the APA's Yankee Scholarship committee a number of years. When me meet to select a recipient I have always encouraged the selection of someone who seemed likely not only to be a good examiner but who also showed promise to do more, perhaps in the research area or in other ways that would advance the field.

In a recent meeting I had with some Board members we discussed how we might get the APA more involved in research. At that time I indicated my view that the APA could not afford to fund research as it is typically thought of; that is, where a grantor provides sufficient funding to support professors and students to complete a project on a given topic. (As you know, $100,000 to do that would be a small amount of funding. The APA could not afford to do this, at least not regularly, if at all.) But, as I indicated above and in this meeting the APA could afford to support graduate students to do thesis (Masters degree) or dissertation (Doctoral degree) research with support from interested university professors. Such students now do a great deal of research guided by professors. The APA must, however, identify professors at research universities and work with them to engage graduate students in projects of their choice but also of interest to the APA. This is not being done.

I didn't get a very positive response from those at this meeting and, unfortunately, I believe that might be the response from the entire Board and maybe even the membership. The primary reason for this is that most, if not all, Board members have no academic working — background or experience. In addition, and in my view, the polygraph examiner community has deliberately isolated itself from the academic world. Maybe this is because academics generally are not favourable to polygraph testing and most of them, if they have an interest at all, hold negative views, e.g., Lykken, Ben Shakhar, etc.. This suggests to me that we need to find ways to overcome this problem. That can't be done without some hard work and, likely, some disappointments. As I see it examiners, including some serving on the Board, don't understand that much of what they write about and publish for their 'community' is of little or no interest to the broader 'scientific' community.

All of that said let me turn to question #3. As I see it then, the most important step that can be taken is not to identify the particular focus of research that needs to be done. Rather, it is to gain involvement from those at universities to do research of interest to polygraphists. If that is initiated we'd find, I'm sure, that professors and students would choose to do research in areas or on problems that are of interest to them; asking such persons to do research that is of no interest to them but is of interest to us simply wouldn't work.

So, turning to your list of topics I'd choose to do first, research that shows us how to differentiate cleanly, if possible, between 'screening' procedures versus those that are devoted to known events. Polygraph testing does not work the same way in both circumstances and it's necessary to deal with this issue. I don't see this on your list. And, unfortunately, it is seldom the case that even examiners note the distinction between the two approaches.

Second, we are now at the point where computerized polygraph instruments allow us to do things with the data we collect that is otherwise not possible. That is not to say, by the way, that at this time I'd advocate the use of "computer scoring" for decision-making. I do think it ought to be used by all examiners as an assist in decision-making. But I'd certainly like to see research on other advances that are possible with computerized polygraph instruments. I'm confident we can make changes here that would be very positive ones. But, again, I need to point out that I am not referring only to the development of or improvements in scoring algorithms.

Finally, another 'problem' not listed by you but which I see is a very important one and one which doesn't get a lot of attention in the literature is what is the examiners' 'skill' in the conduct of an examination and how is this reflected, if it is, in the polygraphic data. It seems to me that much of what examiners write about — and seem to believe—is that differences between examiners are less important than the 'technique' that is used. By technique many examiners—and the APA—usually mean 'format' without considering the difference. My research shows that 'format' does not account for differences of significance. There are many, usually unmentioned, issues of greater importance; examiner differences being one of them.

Frank Horvath, Ph. D.

Book review

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

DOI: 10.2478/EP-2022-0012

Л. Удалова, С. Чернявський, Д. Алєксєєвої-Процюк, *Поліграфологія: основні терміни і поняття*, Київ 2022. [L. Udalova, S. Chernyavs'kyy, D. Alyeksyeyeva-Protsyuk, *Polygraphy: basic terms and concepts*, Kyiv 2022]

The dictionary presents most frequent basic terms and concepts found in scientific literature and international standards of practice of polygraph examiners.

The authors claim that social and political processes, as well as Russia's military aggression against Ukraine make polygraph examinations increasingly relevant for various state and commercial structures. However, despite the rapid development of polygraphy in Ukraine, the terminology is inconsistent between organisations and even individual departments, which complicates professional activity.

That is why they have arrived with the dictionary providing up-to-date systematised reference information on the basic terms and concepts of polygraphy (locally called “polygraphology”), based on leading world experience and current international and domestic professional sources, to be used by polygraph experts.

To eliminate differences in the interpretation of certain terms, reputable scientists and practicing polygraph experts from the National Academy of Internal Affairs, the State Border Guard Service of Ukraine, the Ministry of Defence of Ukraine, the National Police of Ukraine, and the Ukrainian Polygraph Collegium NGO participated in the writing of the dictionary, thus providing an important step towards reaching a consensus among specialists. The dictionary was also recommended for publication by the Academic Council of Ukraine’s National Academy of Internal Affairs.

The current edition differs significantly from ordinary dictionaries. Rather than in alphabetical order, the terms and concepts are listed according to methodological and thematic interconnections. Such an approach allows to reveal important theoretical and methodological foundations. Some terms are presented with alternative interpretations based on international and APA standards of practice. An important advantage is the use of diagrams and pictures to clarify complex terms and illustrate their interrelations. The material is structured in a logical sequence like a regular manual and can be read from cover to cover.

The whole ends in an alphabetical list of terms with links to specific pages that will help a reader to find clear answers to many professional questions.

Vitalii Shapovalov

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

The Basic Information for Authors

To publication will be accepts unpublished research papers as well as review article, case reports, book reviews and reports connected with polygraph examinations.

Submitted manuscripts must be written in English.

All papers are assessed by referees (usually from Editorial Board), and after a positive opinion are published.

Texts for publication should be submitted in the form of normalized printout (1800 characters per page). Use ScholarOne Manuscripts (for online submission and manuscript tracking.

To submit your manuscript, you need the following files:

– Your manuscript (including a title page with the names of all authors and co-authors);
– A main document file with abstract, keywords, main text and references, which should be kept anonymous if the journal you are submitting to uses double-blind peer review;
– Figure files;
– Table files;
– Any extra files such as supplemental material or biographical notes.

The total length of research papers and review article should not exceed 12 pages, case reports – 6 pages, and other texts (book review, report) – 5 pages.

The first page of paper should contain: the title, the full name of the author (authors), the name of institution where the paper was written, the town and country.

Figures should be submitted both in printed form (laser print, the best) and electronic form.

Tables should be numbered in Roman numerals and figures in Arabic ones.

Figures, tables, titles of figures and titles of tables should be included on a separate page. The places in the text where they are to be included should be indicated.

The references should be arranged in the alphabetical order according to the surnames of the authors.

The references should be after the text.

Each reference should include: the surname (surnames) of the author (authors), the first letter of author's first name, the title of the book, year and place of the publication, the name of publisher, or the title of the paper, the full title of the journal, the year, the volume, the number and the first page of the paper.

For example (in references):

Reid, J., Inbau, F. (1966), *Truth and Deception: the Polygraph ("Lie-detector") Techniques*, Baltimore: Williams & Wilkins.

Abrams, S. (1973), Polygraph Validity and Reliability – a Review, *Journal of Forensic Sciences*, 18, 4, 313.

and (Reid, Inbau, 1966), (Abrams, 1973) inside text.

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

***European Polygraph* use ScholarOne Manuscripts for online submission and manuscript tracking**

https://mc.manuscriptcentral.com/polygraph

Preparing your files

To submit your manuscript, you need the following files:

- Your manuscript (including a title page with the names of all authors and co-authors)
- A main document file with abstract, keywords, main text and references, which should be kept anonymous if the journal you are submitting to uses double-blind peer review
- Figure files
- Table files
- Any extra files such as supplemental material or biographical notes

Step – by – step Instruction for Authors

Step 1: Type, Title, & Abstract

Select your manuscript type. Enter your title, running head, and abstract into the appropriate boxes below.

Step 2: File Upload

Upload as many files as needed for your manuscript in groups of five or fewer. If you have more than five files for your manuscript, upload the first five and then you will have the option to upload an additional five files. This process will continue until ALL files have been uploaded. These files will be combined into a single PDF document for the peer review process.

If you are submitting a revision, please include only the latest set of files. If you have updated a file, please delete the original version and upload the revised file. To designate the order in which your files appear, use the dropdowns in the „order" column below. View your uploaded files by clicking on HTML or PDF.

Your text and figure file(s) will be converted into HTML so that they can be easily viewed with a browser on the Internet. They will also be converted into a .PDF document so that they can be viewed and printed with Adobe Acrobat Reader. The files in the .PDF document will be presented in the order specified.

Step 3: Attributes

You may enter your manuscript attributes/keywords in two different ways. Search for a specific term by typing it into the search box or select your keywords directly from the full list (Ctrl + click for multiple words) and click "Add".

Step 4: Authors & Institutions

Enter your co-authors' information by searching on each of their email addresses below. If they have an existing account, their information can be easily imported to your submission. If necessary, you may add a co-author as a new user in our system by clicking "Create New Author".

Step 5: Reviewers

To suggest a reviewer or request the exclusion of a reviewer, click the Add Reviewer button below and enter their information along with the desired designation.

Step 6: Details & Comments

Enter or paste your cover letter text into the "Cover Letter" box below. If you would like to attach a file containing your cover letter, click the "Select File" button, locate your file, and click "Attach File." Answer any remaining questions appropriately.

Step 7: Review & Submit

Review the information below for accuracy and make changes as needed. After reviewing the manuscript proofs at the foot of this page, you MUST CLICK 'SUBMIT' to complete your submission.

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

Reviewers

Articles submitted to *European Polygraph* in 2022 were subjected to peer review by:

Tuvia Amsel

Marcin Gołaszewski

Frank Horvath

Donald Krapohl

Marek Leśniak

Joanna Ulatowska

EUROPEAN
POLYGRAPH
PUBLISHED SEMI-ANNUALLY
2022 VOLUME 16 NUMBER 2 (56)

Rules and Regulations Concerning Publishing Papers in *European Polygraph*

1. All papers sent to *European Polygraph* by their respective authors undergo preliminary assessment by the Editor-in-Chief.

2. The initial assessment results in the decision whether to send the work for an independent review or return it to the author with the information that it will be not published.

3. Two independent reviewers for "internal reviews" are appointed by the Editor-in-Chief or by the Deputy Editor following consultation with the Editor-in-Chief.

4. The following cannot be independent reviewers: Editor-in-Chief, Deputy Editor-in-Chief, employees of Andrzej Frycz Modrzewski Krakow University, and people with papers published in the issue containing the reviewed paper.

5. The internal review should answer the question whether the reviewed paper is fit for printing and whether it requires any amendments, and if it does, state what they are, and must be in written form, and conclude in an unequivocal verdict concerning publication or rejection of an article.

6. If one of the reviewers provides comments and amendments, but does not disqualify the paper, the Editor pass the comments on to the author, asking for the author's opinion and any amendments.

7. Should the opinions of the author and reviewer diverge, the decision to print the paper or otherwise is made by the Editor.

8. In the case mentioned in 7 above, before making their decision, Editor-in-Chief can appoint another independent reviewer.

9. In exceptional cases, when there are significant circumstances justifying such a decision, and the Editors do not agree with the opinion of the reviewer, Editors may decide to publish a paper against the opinion of the reviewer.

10. The names of reviewers is not disclosed to the author, and the names of authors are not disclosed to reviewers.

11. Book reviews and reports are not reviewed, the decision to publish them is the responsibility of the Editors.