



EUROPEAN

POLYGRAPH

PUBLISHED QUARTERLY

Volume 9

2015

Number 4 (34)



Andrzej Frycz Modrzewski Krakow University



EUROPEAN
POLYGRAPH

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.

Editorial Board

Don Grubin (Newcastle, United Kingdom)
Charles R. Honts (Boise, United States)
Frank S. Horvath (East Lansing, United States)
Donald Krapohl (Fort Jackson, United States)
Genrikas Nedveckis (Vilnius, Lithuania)
Jerzy Pobocho (Szczecin, Poland)
David C. Raskin (Homer AK, USA)
Polona Selič (Ljubljana, Slovenia)
Tuvia Shurany (Jerusalem, Israel)
Igor Usikov (Kiev, Ukraine)
Jennifer M.C. Vendemia (Columbia, United States)
Jan Widacki (Krakow, Poland)
Daniel T. Wilcox (Birmingham, United Kingdom)

Editor-in-Chief

Jan Widacki

Managing Editor

Margerita Krasnowolska

Office

ul. Herlinga-Grudzińskiego 1
30-705 Kraków
mail: m.krasnowolska@gmail.com
oleg1998@gmail.com
www.polygraph.pl

Language Editor

George Lisowski

Statistics Editor

Tadeusz Stanisiz

Cover design

Joanna Sroka, Oleg Aleksejczuk



Publisher Council

of the Andrzej Frycz Modrzewski

Krakow University

Klemens Budzowski
Maria Kapiszewska
Zbigniew Maciąg
Jacek M. Majchrowski

Copyright© by Andrzej Frycz Modrzewski

Krakow University

Krakow 2015

ksiegarnia@kte.pl

ISSN 1898-5238

Circulation 200 copies



Contents

- Donald J. Krapohl, Walt Goodson: *Decision Accuracy for the Relevant-Irrelevant Screening Test: Influence of an Algorithm on Human Decision-Making* 189
- Jan Widacki: *Discoverers of the Galvanic Skin Response* 209

Report

- Karolina Dukała: *Report from a national scientific conference on the Instrumental and Non-instrumental Methods of Detection of Deception – Current Courses in Investigations and Legal Framework (Instrumentalne i nieinstrumentalne metody detekcji nieszczerości – aktualne kierunki badań i uwarunkowania prawne)* 223

Book reviews

- O.M. Ribalchenko, V.V. Vasilenko, I.M. Kozub: *Zbirnik testiv poligrafnoj pierievirki [Ukrainian title], Zbornik testov poligrafnoj proverki [Russian title] (Collection of polygraph tests), Wydawniczyi Budinok Melitopolskoi Miskoi Drukarni, Melitopol 2014* 229

■ V.A. Varlamow, G.V. Varlamow, Z.Y. Polovnikova: <i>Trening po razshifrovke poligramm. Metodicheskoe posobie, [in Russian], (Training in reviewing polygrams), SPD Moliar S.V., Kyiv 2010</i>	230
■ L.G. Aleksyeev: <i>Psichofiziologia detektsyi lzhi, Metodologiya [in Russian] (Psychophysiology of lie-detection. Methodology), Masterskaya Prikladnoi Psichofiziologii, Moscow 2011</i>	230
<hr/>	
■ The Basic Information for Authors	233
<hr/>	
■ Rules and regulations concerning publishing papers in European Polygraph	235
<hr/>	
■ Reviewers	237
<hr/>	
■ Ordering Information.....	239



Donald J. Krapohl
Walt Goodson
American Polygraph Association
United States

Decision Accuracy for the Relevant-Irrelevant Screening Test: Influence of an Algorithm on Human Decision-Making¹

Key words: Relevant-Irrelevant Screening Test, Relevant-Irrelevant Test, Accuracy of R-I list, Screening, decision-making

The Relevant-Irrelevant (RI) has been used as a polygraph screening technique by several decades longer than any other. It has demonstrated practical value in prompt-

¹ This is the third in a series of research articles on the RI screening test. The authors would like to express appreciation to the following examiners for participating in this study: Gus Trevino, Derick Walker, Timothy Upright, Daniel Scheel, Jerry Lesikar, Bill Horton, Steven Davis, Eddie Hutchinson, Dana Wickland, Leo Perez, and Eduardo Garza. We also are grateful to Mark Handler for his helpful suggestions to an earlier version of this paper. The first author is a Past President of the American Polygraph Association, and author of the Elsevier textbook *Fundamentals of Polygraph Practice*. The second author is President of the American Polygraph Association, and contributing author to past articles in this publication. The views expressed are the authors' own, and do not necessarily represent those of the Department of Defense, US Government, or the Texas Department of Public Safety. Comments can be sent to Krapohld@gmail.com.

ing self-report from applicants and employees of behaviours of interest to employers. The RI has certain strengths that have made it an attractive alternative (Krapohl & Shaw 2015). With no comparison questions, the RI is not subject to criticisms that the examiner must manipulate the examinee in some fashion to make the technique effective, as probable-lie comparison question techniques may. It is more flexible than most other methods, accommodating from two to five relevant questions in a single test series. The RI may also be more resistant to countermeasures, at least of the type in which examinees induce reactions to comparison questions.

These advantages have kept the RI in widespread practice well after the development and validation of other polygraph screening techniques. However, the existence of these other techniques has brought about comparisons that have not always favoured the RI. For example, the RI does not seem to be amenable to any effective form of manual scoring, a standard for every other screening technique. As such, interpretation of RI data is far more susceptible to individual differences among evaluators. Lenient evaluators can make conclusions highly different from those of conservative evaluators because decisions are based on global impressions of the data, impressions that resist quantification. This was recently demonstrated in a compelling manner in a study where pairs of four experienced polygraph examiners produced opposite opinions in 51 of 100 field RI cases (Krapohl & Rosales 2014).

Another challenge to the RI has been in the finding that its decision accuracy does not compete well against other techniques: Other validated screening techniques have shown higher decision accuracy. This should not be entirely surprising given that a test's reliability sets the limit for its validity, and the RI's unimpressive reliability *should* naturally translate to a lower accuracy. To determine whether RI could possibly achieve accuracy comparable to other screening techniques, it seems reasonable as a first step to try and resolve RI's reliability problem. This would require some form of data quantification.

Past efforts to develop a manual scoring system for RI charts have not produced satisfactory results (Ansley & Weir 1976, Krapohl, Senter & Stern 2005). With computer polygraphs now standard across the profession, an obvious alternative to manual scoring could be the use of an automated algorithm. How automated analysis might improve reliability is self-evident: software analyses the data the same way each time, and is unaffected by human foibles such as bias, mood, attitudes, fatigue, limited experience, and/or poor training. Algorithms are far more reliable than humans in managing complex data needed to form decisions. It seems reasonable, therefore, that an algorithmic approach could increase reliability of data interpretation, and set the stage for potentially greater decision accuracy. a prototype of an RI algorithm

was developed in the early 2000s (Harris & McQuarrie ca 2001). Its unpublished preliminary results were promising, but the algorithm was never subject of an independent assessment, nor was it released to the field. Its influence on decision-making was, therefore, untested.

Algorithms are not without shortcomings. What polygraph algorithms may do less well is identification of artefacts and assessment of stability in the data, an area where humans may have an advantage. Humans are fairly proficient at pattern recognition, a skill helpful in establishing context for the assessment of individual responses, and they can adjust their decision process accordingly. Humans may also be better at detecting idiosyncratic tell-tale patterns for individual examinees in a way that allows a more informed decision than would arise from a strictly statistical one based on group averages.

Based on the unique and disparate capabilities of humans and machines, it is our hypothesis that the combination of human evaluation and algorithmic analysis may improve decision accuracy and interrater reliability for the RI over either human or machine alone. There is no published evidence to suggest whether this is a valid hypothesis, as the polygraph literature is totally silent on human-machine decision making. Our expectation of better accuracy using human-machine collaboration is based on findings from a totally different field, that of weather prediction.

Weather forecasting is heavily relied on in a range of endeavours such as agriculture, travel, construction, civil defence, and others. Consequently, a great deal of attention and funding, has been directed toward developing and assessing weather models, and the place for human decision-making in the overall process. For example, a study of the combination of statistical decision-making with human refinement in weather prediction was undertaken by Carter and Polger (1986). In their 20-year assessment of historical weather data, Carter and Polger found that the most accurate prediction of weather events came when forecasters modified the statistical prediction with factors they knew were influenced by their locality. The National Weather Service models provided the larger framework, and meteorologists' personal experience and knowledge modified those predictions to produce the highest accuracy for their regions of the country. In his book *The Signal and the Noise* (2012), Nate Silver noted that the contribution of human input to computer weather forecasts has been relatively flat at about 25% for precipitation and 10% for temperature, even as forecasting and computer modelling has improved substantially over the decades. The weather modelling literature makes a strong case in that domain that human adjustments of statistical predictions are better than either method alone.

Polygraph decision-making, like weather prediction, is a combination of statistical averages and local variations. In other words, polygraph examiners apply standardised scoring methods to their data, but may consider factors concerning each individual that may adjust their decisions. As an example, if the polygraph data indicate deception, but an examiner observes that his examinee has become emotionally distraught during testing, a statistical decision of deception may be suspended. Examiners may also informally or unconsciously consider extra-polygraphic information, adjusting their scores in the direction they judge most likely correct (Elaad, Ginton, & Ben-Shakhar 1994), which can move decisions to or from inconclusive when the physiological data are ambiguous. Barland (1988) also suggests that an examiner's knowledge and appreciation of base rates may make them more cautious in ascribing deceptive intent when base rates are low.

Though the polygraph literature shows some research that compares human decision accuracy against that of an algorithm (Blackwell 1999, Krapohl & McManus 1999) we did not find any on the influence of algorithm outputs on the validity and reliability of human polygraph decisions. Moreover, the literature is completely silent on the melding together of human and statistical assessments to form polygraph decisions. Our two research questions were therefore straightforward: what is the influence of algorithmic results on human decision-making, and; can validity and reliability be improved by integrating human and algorithmic results to form a single polygraph result? We took advantage of access to skilled polygraph examiners in a large state agency, a substantial archive of verified field RI screening cases, and the results of a prototype RI algorithm to test these questions.

Method

Cases

In the late 1990s, the US government sponsored a research study on the RI screening test using job applicants for security positions at a large metropolitan airport. Confirmation of ground truth was established by record checks and urinalysis. There were four relevant topics covered in those cases: convictions or fines for traffic violations in the State of Georgia in the previous seven years, having been granted bankruptcy in the previous seven years in the State of Georgia, having used marijuana in the previous 30 days, and having been convicted of a felony in the State of Georgia. a description of the original RI research study that produced this archive of cases can be found in Krapohl, Senter, and Stern (2005), and is not recounted here.

We randomly chose electronic polygraph charts of 50 confirmed deceptive cases in which the examinee had been deceptive solely to the question regarding marijuana use, and another 50 from cases in which the examinees had been truthful to all of the relevant questions. The rationale for this sampling approach was fully explained in Krapohl and Rosales (2014), but in brief, it was based on minimum criteria for the quantity and fidelity of the data. The use of cases where there was deception only to a single question was based on earlier findings from Krapohl, Senter, and Stern (2005) that decision accuracy was higher when there were multiple deceptions in a single case, but that multiple-deceptions represented a small minority of all cases. We concluded that the use of cases with multiple deceptions in the sample would limit generalisability of the results, and we eliminated those cases. There was no attempt to exclude single-deception cases that had been selected in previous studies.

All cases had three charts and at least four presentations of each of the four relevant questions. In the original file some cases had four charts, the last being called a “clearing chart.” Clearing charts are used in cases where the testing examiner determined that there were indications of deception on a question in the first three charts, and the examiner used the fourth chart solely to “clear” the remaining questions. Because the presence or absence of a clearing chart would indicate the opinion of the original examiner, those charts were not made available to the blind scorers in this study. As a result, all scorers saw only the first three charts of each RI case.

The physiological data were converted to PDFs, with one file per case, and randomly numbered from 1 to 100 for the first phase of the study, and re-randomised and numbered from 101 to 200 for the second phase.

Blind Evaluators

There were 11 volunteer evaluators from the Texas Department of Public Safety. All held polygraph licenses in the State of Texas. Their average field experience was 6.1 years (ranging from 1 to 20), and their average annual polygraph-related training was 82 hours. All had received training in RI screening techniques and global test data analysis. Four of the examiners had experience conducting RI cases in the field. The examiners all averaged 12 screening examinations in the field per month and all of these examinations were vetted through a 100 percent quality assurance programme.

Algorithm

Johns Hopkins University Applied Physics Laboratory developed the prototype algorithm using the cases collected in the RI project (Harris & McQuarrie ca 2001.) The features used in the algorithm relied on derivatives and weighting of reordered data

points for which it would be difficult or impossible to map to traditional physiological features (e.g., the absolute derivative of reordered abdominal respiration data between the Xth and XXth percentile in time window X to XX seconds and a proportionate weighting of X%). The details of the algorithm features are beyond the scope of this project, and the algorithm itself is proprietary, so the features will not be described further. Decision accuracy of that algorithm in a previous study indicated a mean of 73% for cases that included single and multiple deceptions (Krapohl, Senter & Stern 2005).

The prototype RI algorithm was run against all the 100 cases, and the results were recorded in a spreadsheet. Artefact management was handled by the algorithm, with the first author making corrections only for erroneous question labels. The algorithm results for each case were provided to the blind evaluators only in the second phase of the study. A decision was reached to consider probabilities of deception rendered by the algorithm as lower than 0.30 as No Significant Responses (NSR), while those above 0.70 were called Significant Responses (SR). All the others were called No Opinion (NO).

Data Analysis

Some blind evaluators reported decisions of Deception Indicated (DI), No Deception Indicated (NDI), and Inconclusive, while others used the equivalent labels used in screening, which are Significant Reactions (SR), No Significant Reactions (NSR) and No Opinion (NO). To permit easier comparison to earlier RI studies with these cases, the decisions are reported here as SR, NSR, and NO, respectively, to be consistent with the labelling convention used in Krapohl and Rosales (2014), and Krapohl, Senter, and Stern (2005).

A spreadsheet was constructed in Microsoft Excel[®], and the decisions were recorded. Decisions were coded as -1 for SR, +1 for NSR, and 0 for NO.

Procedure

The study was conducted in two phases. In Phase 1, examiners were told to evaluate individually the RI charts using global analysis in the manner they were accustomed, and to record their decisions² on a data sheet. The examiners were instructed to make decisions by case, not by individual test question. Examiners were not informed of

² The RI is typically the first in a multi-step screening process, wherein all reactions to RI test questions are resolved by further interviewing and more focused testing. Decisions of SR or NO are not appropriate to RI screening alone, but the labels have been used here for convenience.

ground truth or of the test questions. They were asked to work independently, and – to avoid sharing their opinions about the cases with other evaluators – they were also given 30 days to complete the task, with extensions if requested. Examiner instructions are found in Appendix A.

Three months after the receipt of all results from the blind evaluators, the procedure was repeated as Phase 2. In that iteration the file numbers were re-randomised and re-labelled, and the algorithm results were posted on the first page of the PDFs with the physiological data. Examiners received the same instructions as in the first phase, with an extra paragraph added in the instructions that read:

Please review the PolyScore results on the first page of the PDF of each case. Previous research has found this algorithm performs as well as a competent examiner in blind evaluation of RI charts. You are not obligated to use the algorithm results, but only to consider them in formulating your own opinion.

Results

Phase 1: Global Analysis

Table 1 lists the average accuracies for the 11 blind evaluators of the 100 RI cases using global analysis. Average decision accuracy was 74.4% for deceptive cases, and 52.7% for truthful cases, for an overall average of 63.6%. Decision accuracy for deceptive cases was significantly greater than chance ($z = 2.51$, $p < 0.05$), but not for truthful cases ($z = 0.27$, ns). When No Opinion decisions were excluded, overall decision accuracy was 66.9%, which was also significant greater than chance ($z = 2.86$, $p < 0.05$). Collectively, the 11 examiners had a higher decision accuracy for deceptive cases than for truthful cases ($z = 2.25$, $p < 0.05$).

Table 1. Average percentages of correct, incorrect, and No Opinion (NO) decisions by ground truth, with and without NO decisions, for 11 blind evaluators of 50 deceptive and 50 truthful RI cases.

	Deceptive	Truthful	Overall	w/o NO
Correct	74.4	52.7	63.6	66.9
Incorrect	23.5	39.5	31.5	33.1
Inconclusive	2.2	7.8	5.0	

For the entire 100-case sample the average SR rate was 56.9%, 38.1% NSR, and 5.0% NO. The rate at which the 11 evaluators made SR, NSR, and No Opinion decisions varied substantially (See: Table 3). With a base rate of 50 deceptive cases for the 100-case sample, evaluators made from 25 to 78 decisions of SR. With the same base rate of truthful cases evaluators made from 22 to 67 NSR decisions. Evaluators made No Opinion decisions in 0% to 13% of the 100 cases.

Table 3. Number of SR, NSR, and No Opinion decisions of the 11 evaluators of 100 RI cases.

	Examiner										
	1	2	3	4	5	6	7	8	9	10	11
SR Decision	52	78	66	25	72	71	61	47	52	46	56
NSR Decision	47	22	34	67	26	23	32	45	35	53	35
NO Decision	1	0	0	8	2	6	7	8	13	1	9

Discussion

Decision accuracy in Phase 1 of this study (63.6%) corresponded well with the findings of Krapohl and Rosales (2014) of 59.3%, and a reanalysis of Krapohl, Senter, and Stern (2006) of single-deception RI cases showing 63.0% decision accuracy. Similarly, the average proportion of paired agreement among evaluators in Phase 1 of 60.8% was highly similar to that of Krapohl and Rosales (2014) at 59.7%. These comparable findings across different data sets using different scorers point to lacklustre performance of global evaluation when it is used alone in RI cases.

Phase 2: Global Analysis + Algorithm

Table 4 lists the average accuracies for the blind evaluators when they had access to algorithm results when conducting their global assessment of the 100 RI cases. Average decision accuracy was 72.5% for deceptive cases, and 60.7% for truthful cases, for an overall average of 66.6% that was significantly greater than chance ($z = 2.38$, $p < 0.05$). As was found in Phase 1, decision accuracy for deceptive cases was significant ($z = 3.27$, $p < 0.05$), but it was not so for truthful cases ($z = 1.52$, ns). When No Opinion decisions were excluded, overall decision accuracy was 75.3%, which was also significantly greater than chance ($z = 4.98$, $p < 0.05$). There was no difference in accuracy between truthful and deceptive cases ($z = 1.77$, ns).

Table 4. Average percentages of correct, incorrect, and No Opinion (NO) decisions by ground truth, with and without NO decisions, for 11 blind evaluators with access to algorithm results of 50 deceptive and 50 truthful RI cases.

	Deceptive	Truthful	Overall	w/o NO
Correct	72.5	60.7	66.6	75.3
Incorrect	18.4	25.2	21.8	24.7
Inconclusive	9.1	14.0	11.5	

By way of comparison, for deceptive cases the algorithm results were 68% correct, incorrect for 16%, and NO for 16%. For truthful cases algorithm results were correct for 72%, incorrect for 14%, and NO in 14% cases. Detection of deceptive ($z = 2.91$, $p < 0.05$) and truthful cases ($z = 3.19$, $p < 0.05$) were above chance, and there was no difference in detection rates between the two types of cases ($z = 0.62$, ns). Overall correct decisions were 70% with NOs included, and 82.3% without NOs.

The percentages of agreement between each evaluator's decisions and ground truth, and between each pair of evaluators is listed in Table 5. Interrater agreement ranged from 44.0% to 88.0%, with a mean of 64.6%. All but one paired-agreements did not exceed chance expectancy. There was unanimous agreement among all evaluators in 18 of the 100 cases, and of those, 15 were decisions of deceptiveness, three for truthfulness. All unanimous decisions were correct. Of the 50 deceptive cases there were 29 in which at least one evaluator made a decision opposite (NSR vs. SR, either way) to the remaining 10 evaluators. For the 50 truthful cases, 39 had at least one opposite decision among the evaluators.

Overall average agreement between individual evaluator decisions and the algorithm decisions was 66.6% ($z = 4.71$, $p < 0.05$). The average was 72.5% for deceptive cases ($z = 4.02$, $p < 0.05$), and 60.7% for truthful cases ($z = 2.74$, $p < 0.05$).

Table 5. Percentage of agreement for decisions on 100 RI cases between ground truth and each examiner, and between pairs of examiners. All percentages were greater than chance ($p < 0.05$) except that marked *. Chance agreement = 33.3%.

	Examiner										
	1	2	3	4	5	6	7	8	9	10	11
Ground Truth	71	68	72	63	66	61	64	64	68	73	63
Examiner 1		67	67	68	55	50	63	66	70	64	63
Examiner 2			66	74	64	56	70	75	80	61	88
Examiner 3				62	67	66	64	60	62	72	67
Examiner 4					52	44*	65	65	69	62	71
Examiner 5						67	56	54	66	69	61
Examiner 6							57	54	56	66	58
Examiner 7								66	67	66	69
Examiner 8									71	59	72
Examiner 9										61	79
Examiner 10											63

In Phase 2, the average SR rate was 48.9%, 39.5% NSR, and 11.5% NO. The rate at which the 11 evaluators made SR, NSR, and No Opinion decisions is listed in Table 6. Among the 50 deceptive cases, evaluators made from 30 to 70 decisions of SR, and from 24 to 63 NSR decisions. Evaluators made No Opinion decisions in 0% to 20% of the 100 cases.

Table 6. Number of SR, NSR, and No Opinion decisions for 100 RI cases by 11 evaluators who viewed algorithm results.

	Examiner										
	1	2	3	4	5	6	7	8	9	10	11
SR Decision	37	45	54	30	70	68	43	43	48	56	44
NSR Decision	63	40	44	48	24	18	37	47	37	38	39
NO Decision	0	15	2	22	6	14	20	10	15	6	17

Discussion

A comparison of Phase 1 and Phase 2 data hints of increases in decision accuracy (63.6% vs 66.6%, respectively) and interrater agreement (60.8% vs 64.6%, respectively), but the differences proved to be statistically insignificant. Given that the al-

gorithm accuracy was higher than examiner decisions that considered the algorithm results suggested the use of automated analysis as decision support may not be its best role in the decision process.

Post Hoc: Algorithm + Examiner

By itself, the algorithm had a decision accuracy higher than the average accuracy of humans even when they had access to the algorithm decisions. We then considered whether a two-step process, wherein the algorithm would be afforded the first assessment of the data would offer a better approach, and only in cases where the algorithm was unable to decide, the decision would be submitted to evaluation by the examiners. Fortunately, this hypothesis could be successfully tested with the existing data.

We substituted algorithm decisions for those of the examiners in Phase 1, where the examiner had not seen the algorithm results. In cases where the algorithm had produced inconclusive results, the examiners' global evaluations were retained. While this approach to decision making would be expected to improve interrater agreement, inasmuch as all but the 15 inconclusive algorithm decisions would be identical, whether decision accuracy would show a benefit was an open question.

Results

The Algorithm + Examiner (A+E) decision rules boosted overall accuracy, the improvement coming almost exclusively from the correct identification of truthful cases. a reduction of overall error rates fell short of significance. The effect of the A+E decision rules had no effect on deceptive cases. Pair agreement increased, and there was a significant reduction in opposite results arising from among pairs of scorers. See Table 7.

As Table 7 shows, decision accuracy for the A+E method was significantly higher than the results by examiners alone. Virtually all of the benefit came from the higher accuracy with truthful cases. There were significant improvements in reliability and opposite decisions as well.

Table 7. Decision accuracy, interrater reliability, and statistical probabilities of the differences for two decision processes for 11 scorers of 100 RI screening cases.

		Original Examiner Decisions (%)	Decisions of A+E (%)	Diff p
Overall	Correct	63.5	79.5	<0.01
	Error	32.5	19.5	0.051
	No Opinion	5.0	1.0	0.10
Deceptive Cases	Correct	74.4	79.8	0.36
	Error	23.5	19.8	0.52
	No Opinion	2.2	0.4	0.26
Truthful Cases	Correct	52.7	79.1	<0.001
	Error	39.5	19.3	<0.01
	No Opinion	7.8	1.6	<0.05
Reliability	Paired Agreement	60.4	93.0	<0.001
	Opposite Decisions	29.8	6.4	<0.001

To determine whether these findings would generalise to a new sample, we reanalysed the decisions of four experienced examiners for 100 RI screening cases in a previous RI study (Krapohl and Rosales, 2012) using the steps described earlier. Table 8 compares the original decisions with those of the new method. There was an increase in overall correct decisions with the A+E method, but it did not achieve statistical significance. Again, most of the improvement appears to be attributable to the better discrimination of truthful cases. There was, however, a significant difference for errors with deceptive cases, where the original examiner's errors were lower than those of the A+E method.

Table 8. Decision accuracy, interrater reliability, and statistical significance of the differences for two decision processes using examiner decisions from Krapohl and Rosales (2012).

	Original Examiner Decisions (%)	Decisions of A+E (%)	Diff p
Overall			
Correct	59.3	72.0	0.06
Error	32.5	26.8	0.38
No Opinion	8.3	1.3	<0.02
Deceptive Cases			
Correct	81.5	74.0	0.20
Error	12.0	25.5	<0.01
No Opinion	6.5	1.3	0.06
Truthful Cases			
Correct	37.0	70.0	<0.001
Error	53.0	28.0	<0.001
No Opinion	10.0	2.0	<0.02
Reliability			
Paired Agreement	60.5	94.3	<0.001
Opposite Decisions	24.7	3.5	<0.001

General Discussion

The findings from the reanalysis of the Krapohl and Rosales (2012) data from four examiners are largely in accord with those from the first sample from 11 examiners (Table 7). Both data sets show marked benefits of the A+E method for interrater agreement and a reduction in opposite decisions. They also share a common finding that the benefit is loaded on the detection of truthful cases.

Where they most noticeably differ is in the error rate for deceptive cases. As far as the decisions of the examiners in Phase 1 of this study (Table 7) are concerned, there was a non-significant reduction in false negative errors for those cases when the A+E rules were imposed, whereas there was a statistically significant increase when the same A+E rules were used with the Krapohl and Rosales (2012) examiners (Table 8).

Because screening tests, as the first step in the successive hurdles approach, must be sensitive to deceptiveness this difference warrants attention.

A possible explanation lies in differences among the groups of examiners used in the present study and those in the Krapohl and Rosales (2012) research. The present group of 11 examiners came from a state agency in which the RI screening test is taught, but not often used. In contrast, the four Krapohl and Rosales (2012) examiners were federal examiners with substantial field experience with the RI screening test, and had at one time all worked for the same federal agency using this method. The difference in experience could be offered as a possible source of the disparate performance with deceptive cases.

The evidence of such an experience difference between the sample groups is not straightforward, however. When Tables 7 and 8 are compared they show that both groups had highly similar overall decision accuracy, proportions of paired agreement, and rates of opposite decisions. It would be difficult to discern which group was the more experienced based solely on their overall performance data.

Another explanation may be that one group had a bias in decision-making that would be manifested in the types of errors the group made. For example, if one group of examiners had a bias away from certain kinds of decisions, and the A+E method did not, going from one method to the other could shift the kinds of errors made. In reassessing the Krapohl and Rosales (2012) data, it appears their lower rate of errors with deceptive cases may be attributable in part to a general reluctance of those examiners to render NSR decisions. Clearly, if examiners avoid NSR decisions, false positive errors should occur less often. A comparison of average NSR rates from the Krapohl and Rosales (2012) group to those of the examiners in Phase 1 of the present study found significantly fewer NSR decisions ($z = 2.14, p < 0.05$) among the former group, as well as fewer NSR decisions than reached by the algorithm ($z = 4.20, p < 0.01$). Three out of the four examiners in the Krapohl and Rosales (2012) group had decision accuracy statistically *lower than chance* with truthful cases. With the algorithm having a dominant effect on decisions in the A+E arrangement, and lacking the bias, it might be anticipated that the Krapohl and Rosales (2012) examiners would have fewer false positive errors without the A+E rules than with them. This may account for the higher false positive error rate in the A+E arrangement versus the decisions of the examiners in the Krapohl and Rosales (2012) study, an effect not observed among the decisions of the 11 examiners in Phase 1. These findings merit future research.

Summary

We found that allowing examiners to make decisions on the RI screening data only when the algorithm produced inconclusive results produced marked improvements in interrater reliability and detection of truthfulness, and a reduction in opposite decisions among examiners. In one comparison we found the A+E method increased overall decision accuracy. In the cross validation sample the improvement missed significance. The types of errors introduced by the A+E method may be different from those examiners make on their own, though overall error rates are similar. Because the algorithm used in this study is not generally available, our accuracy findings in the A+E arrangement are not expected to represent those of the RI screening test as currently practiced in the field. The converging evidence for decision accuracy for the RI screening test using only global analysis is about 62%, with a similar proportion of agreement among independent examiners. Figure 1 summarises the reliability and accuracy findings of the decision data collected in this study.

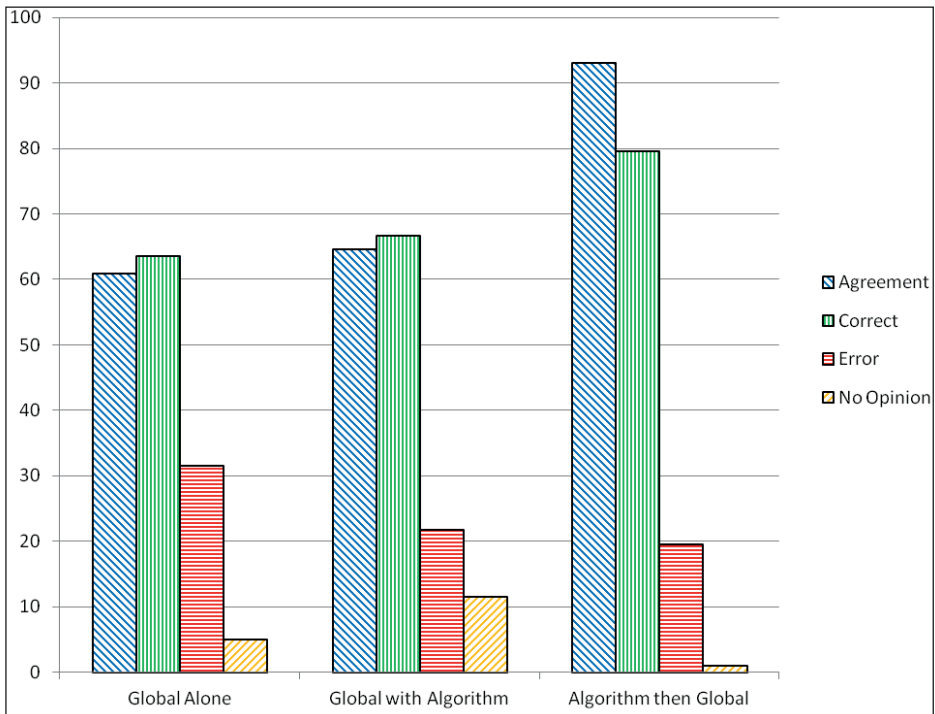


Fig. 1. Summary graph of the three approaches in this study for analysis of RI poly-graph charts.

References

- Ansley N., Weir R. (1976): a numerical scoring system for Relevant-Irrelevant polygraph tests. Paper presented at the 1976 Annual Seminar of the American Polygraph Association.
- Barland G.H. (1988): The polygraph test in the USA and elsewhere. In A. Gale (Ed.) *The polygraph test: Lies, truth and science*. Sage Publications, London.
- Blackwell N.J. (1999): Polyscore 3.3 and psychophysiological detection of deception examiner rates of accuracy when scoring examinations from actual criminal investigations. *Polygraph* 28 (2), 149–175.
- Carter G., Polger P. (1986): a 20-year summary of National Weather Service verification results for temperature and precipitation. Technical Memorandum NWS FCST 31. Washington DC, National Oceanic and Atmospheric Administration.
- Elaad E., Ginton A., Ben-Shakhar G. (1994): The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, 7 (4), 279–292.
- Harris J.C., McQuarrie A.D. (ca 2001): The Relevant/Irrelevant Algorithm Description and Validation Results. The Johns Hopkins University Applied Physics Laboratory.
- Krapohl D., McManus B. (1999): An objective method for manually scoring polygraph data. *Polygraph*, 28 (3), 209–222.
- Krapohl D., Rosales T. (2014): Decision accuracy for the Relevant-Irrelevant Screening Test: a partial replication. *Polygraph*, 41 (1), 20–29.
- Krapohl D., Senter S., Stern B. (2005): An exploration of methods for the analysis of multiple-issue relevant/irrelevant screening data. *Polygraph*, 34 (1), 47–61.
- Krapohl D.J., Shaw P.K. (2015): *Polygraph Screening*. In *Fundamentals of Polygraph Practice*. Academic Press, San Diego, CA.
- Silver N. (2012): *The Signal and the Noise: Why so Many Predictions Fail – but Some Don't*. Penguin Books, New York.

Appendix A

Examiner Instructions

Thank you for volunteering to participate in this project. The data provided by you and your peers will help us in the development of best practices in polygraph screening in general, and the Relevant-Irrelevant (RI) Screening Test in particular. All phases of the project will be completed six months, and we hope to have initial analysis ready for dissemination by summertime. When the analysis is complete, we will issue a formal report of the findings.

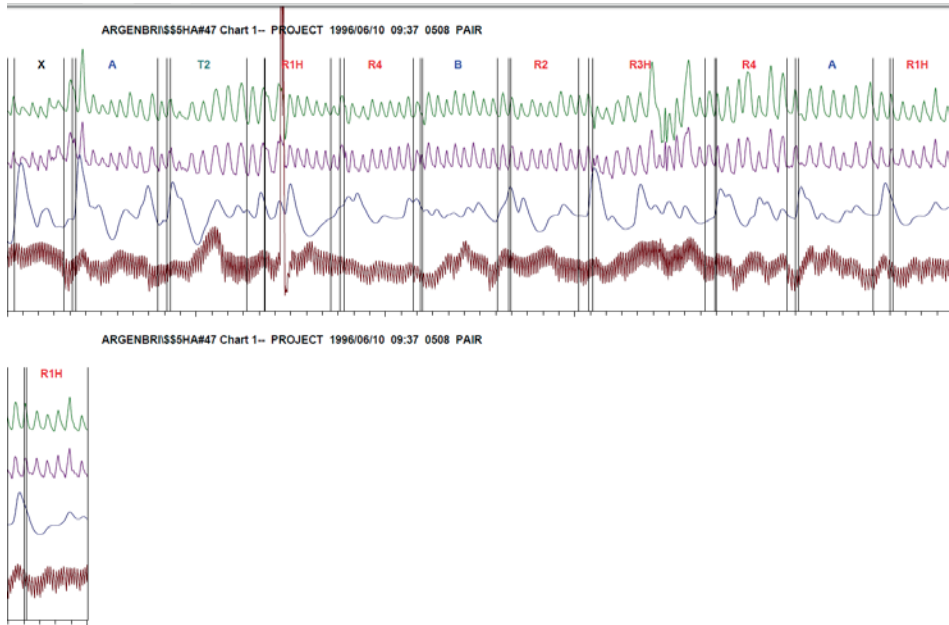
Background

This study has two parts. In the first part you are asked to analyze 100 sets of RI charts collected in a field study in the 1990s. Ground truth was established for each of the relevant issues using record checks, urinalysis, and examinee statements. Some, none, or all of the cases come from deceptive examinees. Your task will be to make a decision for each of the 100 cases of Significant Responses (SR), No Significant Responses (NSR) or No Opinion (NO.)

In about three months we will ask you to look at another sample of RI cases in which we will change the instructions of what you need to do. You will make SR, NSR, and NO decisions on those cases, too. When this phase is completed we will compare the accuracy and reliability of your decisions between the two different parts of the study.

Details

Each of the RI cases is in an individual PDF file, one chart per page. Sometimes the charts were too long to print in a single line, and so they may continue in a different line on the same page. Some portions are repeated between the first and second line. See below as an example.



All cases have three charts, each on a different page. The PDF format will allow you to enlarge or reduce the size of the tracings to your preference.

All relevant questions are marked first with an R, then with a number. In most cases they will be R1, R2, R3 and R4. Sometimes examinees made admissions that required modifications of the relevant questions. The modified relevant questions have the letter H at the end of the label, e.g., R1H, R4H, etc.

Irrelevant questions are denoted by a single letter, e.g., A, B, and C. All other labels indicate technical questions, such as T2, R*C2, and I*C2, which are designed to ensure the examinee is capable of responding. They are not comparison questions, and should not be used in that manner.

In virtually every case there are four presentations of each relevant question. Some have only three, and others may have as many as five.

On the Decision Sheet are the numbers 1-100, each of the numbers representing one of the RI cases in the study. The case numbers are in the PDF file label. You should ignore the case notation in the charts themselves. Simply write your decision for the 100 cases as SR, NSR or NO.

Your Rights

This study relies on volunteers. You do not have to volunteer, and if you do volunteer, you have no obligation to finish the study.

You also are ensured confidentiality. Volunteers will be assigned a number that they will use on the score sheets that only you and I (Don Krapohl) will know. When the study is completed, I will send you, and only you, your individual results. There will be no reports that show which data came from which volunteer. Data will be anonymised or aggregated in the report submitted for publication.

The benefit to you for participating in this study is that you will receive feedback on your accuracy and reliability in the evaluation of RI screening cases. You will also know that your data were important in the development of best practices in polygraph screening.

Your Responsibilities

Our research project requires each examiner to work independently of the other examiners. This means we ask that you not share any information with other volunteers for six months, or when the study is completed. After that time you will be free to discuss whatever you wish without risking the study data.

Also, please remember that accuracy is more important than speed. We have designated 30 days for the volunteers to complete the analysis of the data, but if you find you need additional time it will be given to you.

Contact Information

If you have any comments or questions, please call me at XXX, or via email at XXX (deleted for this publication).



Jan Widacki*

Andrzej Frycz Modrzewski Kraków University
Kraków, Poland

Discoverers of the Galvanic Skin Response

Key words: galvanic skin response, electrodermal activity, Féré, Tarchanoff

Galvanic Skin Response and its role in the instrumental detection of deception

The GSR (galvanic skin response, or electrodermal activity) channel is considered the most diagnostic recording in polygraph examinations, and the best discriminator between people providing deceptive answers to test questions and non-deceptive subjects.

Slowik and Buckley as well as Widacki note that several laboratory studies and experiments in which students played subjects in fictitious crime situations indicated that the galvanic skin response is the most reliable indicator of deception (Slowik & Buckley 1975, Widacki 1977), even though at this time many polygraph examiners with experience in real-life examinations stated that the most reliable indicator of deception is respiration. The GSR is a better indicator in experimental cases as

* jan.widacki@gmail.com

well as in the card test (Reid & Inbau 1966: 256). Yet in the next edition of their book, Reid and Inbau admitted that the GSR is the best indicator in POT and in the Silent Answer Test (SAT) (Reid & Inbau 1977: 287–278). In an ingenious study conducted by Slowik and Buckley cited above consisting of 30 cases, the examiners asked in ‘blind interpretation’ and identified deception with an average accuracy of 87.2% when all three physiological indicators (respiration, GSR, cardio) were analysed. When they based their opinion on respiration alone, the accuracy dropped to 80.5%, to 80.0% when relying on GSR alone, and to 77.1% when working with blood pressure results alone. Thus the results based on the GSR (80.0%) and on respiration (80.5%) were almost the same in a relatively small group (30 people).

In the material covering 36 cases of polygraph examinations conducted in criminal cases in the Department of Criminalistics of the University of Silesia in the late 1970s, whose results were corroborated with valid final court judgements, reactions considered symptomatic were in most cases recorded on the GSR channel. Reactions on that channel were present after critical questions in nearly 70% of cases, while reactions of this type were present on the respiration record in only 51% of cases, and on the cardiovascular record – in only 44% of cases (Widacki 1982: 61, 64).

Similarly, studies of Raskin and collaborators suggested GSR as the most effective parameter (Raskin et al. 1978). For objectivity’s sake, it should be noted that some authors find GSR the least effective parameter (see: Matte & Reuss 1989, 1992). Franz, who resorted to a computer analysis of 100 confirmed field cases, found the electrodermal response to be the most accurate (Franz 1990). According to another author, R. Ryan (Ryan 1989), the GSR is most effective in identifying the deceptive (‘guilty’) subject but it is respiration that allows one to identify the non-deceptive (‘innocent’) subjects most effectively.

In the polygraph examination techniques most frequently applied in recent times, for example, the Utah Zone Comparison Technique (UZCT) in the numerical assessment of the recordings, as for example Empirical Scoring System (ESS) the diagnostic value of the GSR is assessed much higher than respiratory symptoms (Krapohl, Shaw 2015).

Changes in the GSR are the most frequently used indicator in various simplified lie detection procedures. One of them makes use of a new device introduced by the Americans in Afghanistan and Iraq for the initial screening of suspects. It is the so-called Preliminary Credibility Assessment Screening System: (PCASS). The PCASS is a handheld computer or personal digital assistant that attempts to measure stress to decide whether a subject is telling the truth. To detect deception, the PCASS uses

external physiological information (GSR and cardiovascular) collected during an interview with a photoplethysmograph (Gordon, Fleisher 2011: 321).

Thus changes in the GSR are still used both for classical polygraph examinations, and for a variety of simplified procedures of instrumental lie detection based on the observation of physiological correlates of emotion.

It is worth to mention that much of experimental research on the detection of deception makes use of a psychogalvanometer alone rather than of a multi-channel polygraph device. For example, David Lykken's first experiments (Lykken, 1959) which gave rise to the Guilty Knowledge Test (GKT) only made use of a psychogalvanometer and not of a polygraph.

Galvanic Skin Response: a short history of discovery

In 1849, Emil du Bois-Reymond in Germany discovered that the human skin is electrically active. In the later years of the 19th century, Romain Vigouroux, a collaborator of the French neurologist Jean Charcot (1825–1893), measured tonic skin resistance levels in various emotionally distressed patient groups when looking for clinical diagnostic signs during studies of hysteria and hypnosis in Charcot's laboratory. He noticed that the electrical resistance of the skin increases on the sedated part of the body of his hysteric patients (Vigouroux 1879).

In the same laboratory, Charles Féré (1852–1907) found that by passing a low electrical current between two electrodes placed on the surface on the skin, one could use a galvanometer to measure momentary decreases in skin resistance in response to a variety of stimuli of various types, including visual and auditory ones (Féré 1888). In this way Féré discovered that the skin becomes a better conductor of electricity in the presence of external stimuli.

In 1890, a Russian physiologist of Georgian origin, Ivan Tarchanoff (Tarchanov, Tarkhanishvili, 1846–1908) discovered that one could measure changes in the electrical potential between two electrodes placed on the surface of the skin, yet unlike in Féré's experiment, without applying an external source of current. Various stimuli result in a change in the electrical potential and make the pointer of the galvanometer move (Tarchanoff 1890).

As Jaffres proved, the Féré phenomenon and Tarchanoff phenomenon have same physiological mechanism and are two ways of observing and measuring the same phenomenon (Jeffres 1928).

Thus Féré and Tarchanoff are the joint discoverers of the galvanic skin response.

It is thanks to these two scientists that we can resort to galvanic skin responses in contemporary instrumental lie detection.

In 1897, a German, Georg Sticker, was the first to suggest the use of a galvanometer (psychogalvanometer) for lie-detection (Troville 1939), and in 1909 Otto Veraguth used electrodermal response for his experiments with word association. More than 85 years ago, John Larson complemented what had previously been a two-channel polygraph with the psychogalvanometer. Since then a psychogalvanometer has been a significant and standard component of all polygraph devices (Abrams 1989: 4), while the Galvanic Skin Response (GSR) is among the prime and most evident and accurate physiological changes used in instrumental detection of deception.

It is generally assumed that the United States is home to the polygraph. Even at the most recent, 50th Seminar of the American Polygraph Association in Chicago, held from 30 August to 4 September 2015, Frank Horvath and Stanley Slowik, in a paper delivered under the title of “The Birthplace of Modern Polygraphy”, tried to convince the audience that the actual birthplace is Chicago. It is true that the United States is the country where the contemporary polygraph was constructed, and which also boasts the longest practice in the use of the machine which began in the 1920s. However, it goes without saying that polygraph examination has European roots. Without these, as well as without European experimental psychology and without the European discoveries in physiology and psychophysiology of the late 19th and early 20th centuries, without European scientists who in some cases (notably Hugo Münsterberg) were also active in the United States, the origin and development of polygraph examinations would not have been feasible. In a nutshell, without European scholars of the order of Angelo Mosso, Vittorio Benussi, Cesare Lombroso, Charles Féré, and Ivan Tarchanoff, the works of John Larson and Leonard Keeler, whom Americans rightly consider to be the ‘fathers of the polygraph examination, would not have been possible (Widacki 2012).

Discoverers

Charles Samson Féré was born in July 1852 and died on 22 April 1907. He was a French physician, whose broad range of scientific interests encompassed medicine, psychology, physiology, sexology, and the phenomena of magnetism and hypnosis. In relation to the last of these topics he collaborated with Alfred Binet, the later inventor of a method for practical measurement of intelligence. To a certain extent

he was also involved in criminology to which he contributed the seminal works: *Degeneration and criminality* (*Dégénérescence et criminalité*) published in 1888 and *La Famille névropathique* (*The neuropathic family*), published in 1894. From 1881 Féré was an assistant to Jean-Martin Charcot (1825–1893). As mentioned above, R. Vigouroux also worked with M. Charcot.

In 1888, Féré published a short report entitled *Note sur les modifications de la résistance électrique sous l'influence des excitations sensorielles et des émotions* (A note on changes in electrical resistance under the impact of sensory stimulation and emotion) in the journal *Comptes rendus des séances de la Société de biologie* (Ser. 9, 40, 5, pp. 217–219).

As mentioned above, in his experiment, Féré attached two electrodes connected in series to a weak source of electricity and to a galvanometer to the forearm of the patient whom he later subjected to a range of sonic, olfactory, and visual stimuli. Such stimuli had the galvanometer pointer moving. Initially, Féré believed this to be the result of friction on dry skin. Yet Jaques-Arsene d'Arsonval (1851–1940), a physicist he collaborated with, realised that the change in conductivity was linked to sweating (d'Arsonval 1888). Thus, what we refer today as the 'Féré phenomenon' should more properly and justly be called 'the phenomenon of Féré – d'Arsonval'.



Fig. 1. Ch.S. Féré



Fig. 2. J.A. d'Arsonval

In his publication from 1890, Ivan Tarchanoff presents the discovery that the galvanometer reacts similarly to stimuli even in the absence of an external source of energy (Tarchanoff 1890). The scientist attached electrodes to two randomly chosen points

on the skin of the patient and, by placing a galvanometer between them, proved a difference in potential between the points. When he subjected the patient to such stimuli as Féré did, the electric potential changed, resulting in the movement of the galvanometer. This phenomenon is known as the ‘Tarchanoff phenomenon’.

Both the methods – that of Féré using only an external source of electricity and that of Tarchanoff using solely the endosomatic electrical potential – study and observe the same phenomenon and the same physiological mechanism (Jeffers 1928). For this reason, we can speak of ‘the Féré – Tarchanoff phenomenon’ being used in polygraph examinations. Today, thanks to the experiments of Darrow from the 1930s, we know that it doesn’t use the changes in skin wetness caused by sweating, as the response in the form of a change in conductivity comes before sweat, but from the action of the sweat glands themselves.

Ivan Tarchanoff (born Ivane Tarchan Mauravov Tarkhnishvili), in fact the only East European academic with a significant contribution to the scientific foundations for polygraph examinations, is moreover an exceedingly interesting figure. He was a great scholar of European format, symbolically combining the nations of Central and Eastern Europe. A Georgian aristocrat, a graduate of a Russian university and later a member of the Russian faculty, a professor of the Military Medico-Surgical Academy in Saint Petersburg, he was an avid experimenter in what is broadly construed as physiology and a promoter of science.



Fig. 3. I. Tarchanoff as professor of the Imperial Military Medico-Surgical Academy (by I. Repin)

Known for his progressive and liberal views, Tarchanoff could not win favour among the Russian authorities. There are reasons to believe that it was under their pressure that he was forced to leave his academic chair at the age of 48, in 1894. Still popular in Russia and portrayed by the best Russian painters, notably Ilia Repin he was also popular in Western Europe and among Polish scientific and cultural circles. Tarchanoff popular lectures in Saint Petersburg received coverage even in the Polish daily press. He published in the best West European scientific magazines, and participated in numerous international congresses. His ties to Poland were especially strong. He married a Jew from Vilnius (Polish: Wilno), Elena (Helena) Antokolska, an artist sculptor brought up in the Polish culture. Moreover the Russian academic was a teacher and friend of Napoleon Cybulski whom he successfully recommended as his assistant in Saint Petersburg to the Chair of Physiology of the Jagiellonian University in Kraków. Later Dean of the Medical Faculty and Rector of the Jagiellonian University, Cybulski was among the greatest Polish academics of the turn of the 19th and early 20th centuries, co-discoverer of adrenaline, and one of the first in the world to have obtained a recording of the electrical changes in the cortex of the brain (Widacki 2015), and he is recognised as the creator of the Polish school of physiology.

The fact that in 1905 Tarchanoff arrived in Kraków, at that time part of Austrian Galicia, most probably with the intention of settling in the vicinity of the city for good, is hardly known to Tarchanoff's biographers and researchers of his work (Tsagareli 2012). He made his home near Kraków, in Nawojowa Góra, and had Napoleon Cybulski as a neighbour.



Fig. 4. Tarchanoff's house in Nawojowa Góra (Poland)

Tarchanoff published his last works in cooperation with Cybulski in Kraków in Galicia in 1905–08. The Annual of the Academy of Arts and Sciences in Kraków (1905/1906) recorded that the Faculty of Mathematics and Natural Sciences ‘at its sessions analysed 75 works, of which it approved 71 for publication’ (Rocznik 1906: 83), of which 13 related to botany, and their number included ‘a preliminary report from Professor Tarchanoff entitled “observations on the radioactive properties of plants” (Rocznik 1905/1906: 85)’. The article was finally published in the French version of the Bulletin of the Academy of Art and Sciences, with I. Moldenhauer as co-author. In 1907, Cybulski and Tarchanoff together published an article in Polish entitled ‘Kilka słów w sprawie jądów w jelicie prawidłowym’ (A few words on toxins in healthy intestine). The article was published in Lemberg (Polish Lwów) in the Polish language in a medical journal Tygodnik Lekarski (Cybulski, Tarchanoff 1907). It was probably Trakhanoff’s last publication as he died in his home in Nawojowa Góra near Kraków on 24 August 1908 (Nawojowa Góra belongs to Rudawa parish).

It is intriguing to note that, according some sources, e.g. the Great Soviet Encyclopedia and even Wikipedia, Tarkhanow died in Saint Petersburg (Great Soviet Encyclopedia [Bolshaya sovetskaya entsiklopediya 1976, Wikipedia as visited in 2014).

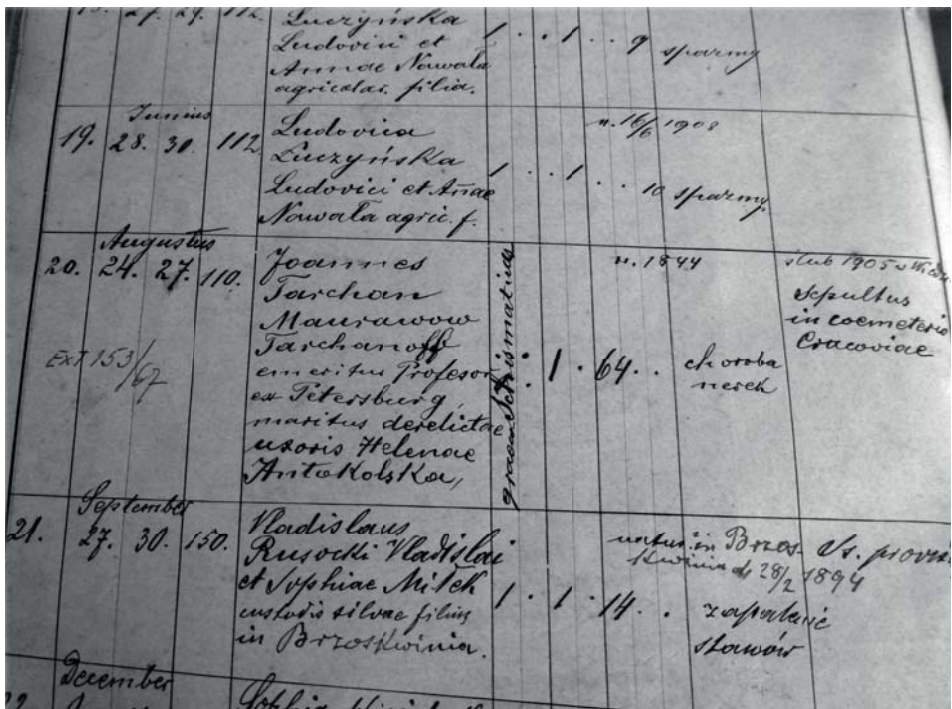


Fig. 5. Rudawa parish register “Liber Mortuorum” (attesting to Tarchanoff’s death)

After his death in 1908, Tarchanoff was first temporarily buried in a Kraków cemetery, yet his body was moved to Saint Petersburg where it was finally interred in the cemetery by the Alexander Nevsky Lavra.



Fig. 6. Tarchanoff's grave in Petersburg

A monument to Ivan Tarchanoff can be found in Tbilisi, on the front of the main building of the Medical University. This is how a co-discoverer of the galvanic skin response symbolically brought together Russians, Georgians, and Poles.



Fig. 7. Monument to Tarchanoff in Tbilisi

References:

- Abrams S. (1989): *The complete polygraph handbook*, Lexington Books, Lexington, Mass., Toronto.
- D'Arsonval A. (1888): Remarques à propos de la production d'électricité chez l'homme, *Comptes Rendus Societe de Biologie*, 40, 142–144.
- Cybulski N., Tarchanow I. (1907): Kilka słów w sprawie jądów w jelicie prawidłowym, *Lwowski Tygodnik Lekarski*, 2 (41), 579–580.
- Bolshaya sovetskaya entsiklopediya (1976): A.M. Prochorow ed., Vol. 25, 3rd edition, Moscow, p. 387. Féré Ch. (1888): Note sur les modifications de la résistance électrique sous l'influence des excitations sensoriels et émotions, *Comptes Rendus Societe de Biologie*, 40, 217–219.
- Franz M.L. (1990): Technical report: relative contributions of physiological recordings to detect deception, DoD contract MDA 904-88-M-6612 Argenbright Polygraph, Inc. Atlanta GA.
- Gordon N.J., Fleisher W.L. (2011): *Effective interviewing and interrogation techniques*, 3rd ed., Elsevier, Academic Press, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Sydney, Tokyo.
- Jeffres L.A. (1928) Galvanic phenomena of the skin, *Journal of Experimental Psychology*, 11 (2), 130–144.
- Krapohl D., Shaw P. (2015): *Fundamentals of polygraph practice*, Elsevier, Academic Press, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Sydney, Tokyo.
- Lykken D.T. (1959): The GSR in detection of guilt, *Journal of Abnormal and Social Psychology*, 43, 6.
- Matte J.A., Reuss R.M. (1989): A field validation study of the Quadri-Zone Comparison Technique, *Polygraph*, 18 (4), 187–202.
- Matte J.A., Reuss R.M. (1992): A study of the relative effectiveness of physiological data in field polygraph examination, *Polygraph*, 21 (1), 1–22.
- Raskin D.C., Barland G.H., Podlesny J.A. (1978): *Validity and reliability of detection of deception*, Washington D.C. National Institute of Law Enforcement and Criminal Justice.

Reid J., Inbau F. (1966): Truth and deception. The polygraph (lie-detection) technique, The Williams & Wilkins Comp., Baltimore.

Reid J. Inabu F. (1977): Truth and Deception. The polygraph (lie-detector) technique, 2nd ed., The Williams & Wilkins Comp., Baltimore.

Ryan R. (1992): Relative validity and utility of examiner diagnosis of truth and deception utilizing respiration, cardiovascular, and galvanic skin response, unpubl. Master's thesis, Reid College of Detection of Deception (quoted from: J.A. Matte: Forensic psychophysiology – using the polygraph, J.A.M. Publications, Williamsville, New York.

Slowik S., Buckley J.P. (1975): Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure and GSR recordings, *Journal of Police Science and Administration*, 3 (3), 305–310.

Tarchanoff I. (1890): Über die die galvanischen Erscheinungen in der Haut des Menschen bei Reizungen der Sinnesorgane und bei verschiedenen Formen der psychischen Thätigkeit, *Pflüg. Arch. Ges. Phys.*, 46, 46–55.

Tarchanow I., Moldenhauer T. (1906): Promienistość indukowana i naturalna roślin i jej prawdopodobne znaczenie dla wzrostu roślin, *Bulletinn Internationale de l'Academie des Sciences de Cracovie Anne 1905, Classe des Sciences mathematiques et Naturelles*, 393–408.

Trovillo P.V. (1939): A history of lie-detection, *The Journal of Criminal Law and Criminology*, 29.

Tsagareli M.G. (2012): Ivane Tarkhanishvili (Tarchanoff). Major Georgian Figure from Russian Physiological School, *Journal of the History of Neuroscience*, 21, 393–408.

Vigouroux R. (1879): Sur le role de la resistance electrique des tissus dans le'electrostatic, *Comptes Rendus Societe de Biologie (Series 6)*, 31, 336–339.

Widacki J. (1977): Wartość diagnostyczna badania poligraficznego i jej znaczenie kryminalistyczne, Wydawnictwo UJ, Kraków.

Widacki J. (1982): Analiza przesłanek diagnozowania w badaniach poligraficznych, Wydawnictwo Uniwersytetu Śląskiego, Katowice.

Widacki J. (2012): European roots of instrumental lie-detection, *European Polygraph*, 6, 4 (22).

Widacki J. (2015 a): Profesor Iwan Tarchanow, *Przegląd Lekarski*, 72/1.

Widacki J. (2015 b): Ivane Tarkhnishvili (Ivan Tarchanoff) and his links with Poland, *Journal of the History of Neurosciences*, 4, 1–9.

Wikipedia, https://en.wikipedia.org/wiki/Ivan_Tarkhanov_%28physiologist%29 [accessed on 05.07.2014].

The article is a part of Polish National Science Centre's (NCN) project no. DEC-2013/11/B/H55/03856

Report



Report from a national scientific conference on the
Instrumental and Non-instrumental Methods of Detection
of Deception – Current Courses in Investigations and
Legal Framework (Instrumentalne i nieinstrumentalne
metody detekcji nieszczerości – aktualne kierunki badań
i uwarunkowania prawne)

A national conference devoted to the methods of deception detection combined with the Congress of the Polish Polygraph Association was held in the building of the College of Law of the University of Law and Public Administration (ULPA/WSPiA Rzeszów–Przemyśl) on 11th and 12th December. Its organiser was WSPiA Rzeszów–Przemyśl, supported by the Polish Polygraph Association, and the Rector of the University, Professor Jerzy Pośluszny, extended honorary patronage over the conference.

The organisers devised the formula of the conference so as to set up a forum for discussing new trends and discoveries in lie detection methods (with a special emphasis on the polygraph) both for the representatives of the scientific world and practitioners. The conference attracted over 70 people dealing with law, psychology, penitentiary sciences, criminalistics, and criminology. Such a solution allowed interdisciplinary discussion on the potential courses of development of methods of deception detection and good practices in polygraph examinations.

The conference was opened by Professor Jerzy Połuszny, Rector of the WSPiA, Professor Czesław Kłak, Director of the College of Law at the WSPiA, and by Marcin Gołaszewski, President of the Polish Polygraphers Association. The keynote lecture by Professor Katarzyna Kaczmarczyk-Kłak (WSPiA Rzeszów–Przemyśl) presented the links between the Constitution of the Republic of Poland and issues in the practical use of the polygraph.

The first panel was devoted to the possibility of using polygraph examinations and interpreting their results in the amended criminal procedure. Professor Jan Widacki (Andrzej Frycz Modrzewski Kraków University) vividly presented the potential consequences and interpretations of the latest statements of the Supreme Court (of 29 January 2015) for polygraph examinations in the Polish criminal procedure. In turn, Professor Czesław Kłak tried to answer the question where and to which stages of the criminal procedure to fit the polygraph on the grounds of the recent amendment of the Criminal Procedure Code, in force since 1 July 2015. The third paper in the panel was presented by Professor Ryszard Jaworski (University of Wrocław), who presented the conclusions drawn on the results of polygraph examinations and discussed the most frequently recurring and potential errors and reinterpretations of the results. The first part of the panel was recapitulated in a lively discussion concerning the current place of the polygraph in the criminal procedure and beyond it, and on the potential uses of the polygraph that have not yet been put into effect.

The panel closed with two papers devoted to the relatively new methods of detection of deception. Karolina Dukała and Dr Romuald Polczyk (Institute of Psychology of the Jagiellonian University) presented the options for training police officers based on the methodology of verbal and non-verbal behaviour analysis, and results of their own studies on the efficiency of use of various methods of detection of deception by police forces. In turn, Marcin Gołaszewski presented in his paper the latest achievements of the world science in lie detection, quoting the most thought-provoking statements from world-class scientists and practitioners gathered in the UK at the first Decepticon: International Conference on Deceptive Behaviour (Cambridge, 24-26 August 2015) and at the Annual Seminar of the American Polygraph Association (Chicago 30 August – 4 September 2015).

The session was continued after the lunch break by Dr Lucjan Wiśniewski (Polish Border Guard), who presented the history of polygraph examinations in Poland before 1990. The question of history continued in the following paper, in which Anna Szuba-Boroń (Andrzej Frycz Modrzewski Kraków University) presented the character and use of the polygraph in contemporary Poland. Thanks to thus designed

plan of presentations, it was easy to observe the changes taking place in Poland in the scope, potential, and purposes of polygraph use. It was fairly easy to notice that the polygraph becomes currently more popular and is applied not only for criminal cases but also, quite intensively, for recruitment to specific posts in uniformed services.

The following presentation concerning the future of the polygraph and its potential uses, especially as a therapy supporting tool. Agnieszka Leszczyńska shared news from the latest evidence-based studies demonstrating efficiency of polygraph use supporting sexual offender therapy following their release from the prison. Although this specific use of the polygraph will require further research and cannot be applied in practice without a very carefully designed system of support, it must be emphasised that the field may be one where polygraph will find application in future. The above is prompted especially by current foreign, especially British, experience.

The first day of the conference closed in two presentations on the use of the GSR channel in polygraph. The first, delivered by Professor Jan Widacki, demonstrated the GSR signal characteristics, and the life and the Polish links of its co-discoverer, Ivan Tarchanoff. The second demonstrated studies conducted by Anna Czapryna (UJ) and Dr Marek Leśniak (University of Silesia) concerning the potential for purposeful distortion of polygraph results by the examinees. This highly interesting paper distinctly explained differences in lie detection accuracy resulting from the employment of various automatic software-based solutions for calculations, and the dangers linked to the use of only one channel – the GSR.

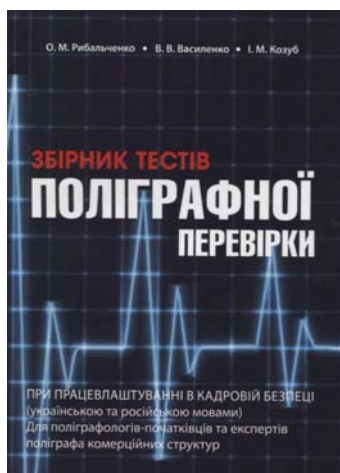
The second day of the session was devoted mostly to the practical use of polygraph examinations. Both issues of difficulties in using the polygraph in the preparation (presented broadly by junior inspector Andrzej Bodzioch (retired)) and court phases of the investigation. The latter was discussed, together with potential solutions offered by the amendment of the Criminal Procedure Code, by a judge, Grzegorz Maciejowski. The possibility of drawing expert conclusions from a polygraph examination in the light of Strasburg standards were presented by Arkadiusz Szajna, followed by Edward Lewandowski who offered a broad, however controversial, discussion of the theoretical aspects of polygraph-based examinations applied in criminalist tactics, and pointed to the need of scientists participating in the investigation of efficiency of various methods of polygraph use. A lively discussion on the justification of non-standard (i.e. not recommended by the American Polygraph Association) polygraph techniques in criminalistic practice followed the panel. The conference closed in a presentation of modern Criminalistic Laboratories of the WSPiA Rzeszów–Przemyśl.

The papers and following discussions followed two clearly visible courses. The first concerned the interdisciplinary nature of the field: currently, global studies on lie detection are conducted with the use of the polygraph and accompanying devices (e.g., infrared cameras, non-verbal behaviour assessment, and eye trackers), which seems to be a promising course in contemporary criminalistic research. The second conclusion was connected to the use of the polygraph in Poland: as far as an increase in the significance of the polygraph in criminal cases and recruitment can be observed, it still seems that, compared to the potential it offers, the polygraph remains too seldom used in practice.

Karolina Dukała*

* dukala.karolina@gmail.com

Book reviews



O.M. Ribalchenko, V.V. Vasilenko, I.M. Kozub: *Zbirnik testiv poligrafnoj pierievirki* [Ukrainian title], *Zbornik testov poligrafnoj proverki* [Russian title] (Collection of polygraph tests), Vydawnichyi Budinok Melitopolskoi Miskoi Drukarni, Melitopol 2014

This small, bilingual book of just 76 pages consists of two identical parts, the first in Ukrainian and the second – with contents identical with the first – in Russian. It is designed for experts running pre-employment procedures for business. Its bibliography contains solely Russian, Belarusian, and Ukrainian literature, as a rule unknown to the Western reader.

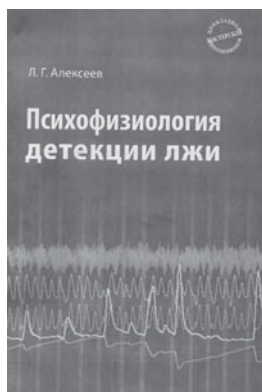
The authors are the deputy president of the College of Ukraine Polygraphers, director of company promising safe staffing, and polygraph examiner – Ribalchenko, and two people (Vasilenko and Kozub) connected to the same company and practical polygraph examinations. Therefore it is justified to believe that the book is also based on their personal experience.



V.A. Varlamov, G.V. Varlamov, Z.Y. Polovnikova: *Training po razshifrovke poligramm. Metodicheskoe posobie*, [in Russian], (Training in reviewing polygrams), SPD Moliar S.V., Kyiv 2010

Training material composed of 29 computer polygraph printouts (polygrams, pp. 6–35). Each printout is accompanied by two questions, one about the psychological traits of the examinee (e.g. the emotional state, personality disorders, etc.) and the other about the test question that was followed by the strongest reaction. The correct answers are provided at the end of the book (pp. 34–36). The contents is training material of high suitability for beginner polygraph examiners, and can be used both for training and testing polygrapher knowledge.

The authors of the publication are competent, their names are known from literature as authors of works on polygraph examinations (Varlamov is a co-author (with V. Knyazev) of a book entitled *Poligraf i ego prakticheskoe promyenyenye* – see: *European Polygraph* 2014, 9, 3 (4), 177) and experienced practitioners.



L.G. Aleksyeev: *Psichofiziologia detektsyi lzhi, Metodologiya* [in Russian] (Psychophysiology of lie-detection. Methodology), Masterskaya Prikladnoi Psichofiziologii, Moscow 2011

Although not distinguished formally, the book consists of three parts. The first, covering chapters from 2 to 6 (pp. 8–38) contains ordered information in psychological

and psychophysiological grounds for polygraph examinations (memory, motivation, physiology of psychological phenomena, the phenomenon of the lie, etc.). Part two, composed of chapters from 7 to 12 (pp. 29–88), discusses polygraph examination as a method of observation and registration of physiological correlates of emotions that, on certain conditions, allows to conclude about deception. The last part, i.e. chapters 13 and 14 (pp. 89–103), is generally devoted, as the author claims, to ‘an actually new technique’, namely the application of the APK Konkord – M ‘hardware and software complex’ (Russian: *apparatno-programmnyi kompleks*). The author suggests that the device is earmarked to be used in customs control on border passes for searches and quick checks. The principle of operation of the Konkord system combines the functions of a traditional polygraph and remote polygraph with analysis of changes in handwriting under the impact of emotion and with speech and video stream analysis. Altogether, the information seems to be exceptionally attractive, yet it is hard to tell whether the method described is routine, and to what degree, or whether it is just a subject of tests and investigation.

On the whole, it seems certain that not only are there plenty of polygraph examinations conducted in the countries of the former USSR whether for business use or for the organs of the state, but also that Russia, Belarus, and Ukraine are home to numerous experimental studies and investigations in the area, which for a variety of reasons are not noted in Western literature.

[J.W.]

The basic information for Authors

To publication will be accepted 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) and in electronic form (diskette, CD), or sent by e-mail to Editorial Office.

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 BASIC INFORMATION FOR AUTHORS

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*, Williams & Wilkins, Baltimore.

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

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

Texts for publication in "European Polygraph" should be mail to:

"European Polygraph"
Andrzej Frycz Modrzewski Krakow University
ul. Gustawa Herlinga-Grudzińskiego 1
30-705 Kraków (Poland)

or e-mail: m.krasnowolska@gmail.com
oleg1998@gmail.com

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.

Reviewers

Articles submitted to *European Polygraph* in 2015 were subjected to peer review by:

Tuvia Amsel

Marcin Gołaszewski

Natan Gordon

Ryszard Jaworski

Jerzy Konieczny

Marek Leśniak

James Allan Matte

Tadeusz Widła

Ordering Information

Please, send your orders by e-mail to ksiegarnia@kte.pl, including:

- full name (first and last in case of natural persons; registered business name in case of legal persons)
- address (permanent address or registered seat),
- address for delivery of your copies of European Polygraph,
- number of successive issues ordered (minimum 4), and number of copies of each issue.
- One year subscription (4 issue): USD 50, € 40.
- Shipment costs is added of the subscription price.
- All subscriptions must be pre-paid to our account:

Krakowskie Towarzystwo Edukacyjne Sp. z o.o. ul. G. Herlinga-Grudzińskiego 1C/
lokC224, 30-705 Kraków

Payment by USD: 24 1020 2892 0000 5102 0222 8161

IBAN PL 24 1020 2892 0000 5102 0222 8161

KOD BIC (SWIFT): BPKOPLPW

Name of the bank: Powszechna Kasa Oszczędności Bank Polski Spółka Akcyjna.

Payment by €: 19 1020 2892 0000 5002 0222 8203

IBAN PL 19 1020 2892 0000 5002 0222 8203

KOD BIC (SWIFT): BPKOPLPW

Name of the bank: Powszechna Kasa Oszczędności Bank Polski Spółka Akcyjna.

All publication dates are subject to change without notice.

Условия подписки

Заказы содержащие имя и фамилию или название компании, учреждения, адрес доставки и номер журнала, от которого должна происходить подписка (минимум 4 номера) необходимо переслать на адрес: ksiegarnia@kte.pl.

Стоимость годовой подписки, охватывающей 4 номера составляет \$ 50 или € 40.

Цена включает в себя стоимость доставки.

Предоплата должна быть произведена на наш банковский счет:

Krakowskie Towarzystwo Edukacyjne Sp. z o.o. ul. G. Herlinga-Grudzińskiego 1C/lokC224,
30-705 Kraków

Payment by USD: 24 1020 2892 0000 5102 0222 8161

IBAN PL 24 1020 2892 0000 5102 0222 8161

KOD BIC (SWIFT): BPKOPLPW

Name of the bank: Powszechna Kasa Oszczędności Bank Polski Spółka Akcyjna.

Payment by €: 19 1020 2892 0000 5002 0222 8203

IBAN PL 19 1020 2892 0000 5002 0222 8203

KOD BIC (SWIFT): BPKOPLPW

Name of the bank: Powszechna Kasa Oszczędności Bank Polski Spółka Akcyjna.

Сроки публикации могут изменяться без предварительного уведомления.