



Tuvya T. Amsel¹

Comparative Review of Polygraph and Other Diagnostic Tools and Methods*

*A lie can travel halfway round the world
while the truth is putting on its shoes*
Attributed to Mark Twain

Abstract

Unlike the polygraph, many diagnostic tools and methods are admissible in court in spite of being as accurate as, or less accurate than, the polygraph. The data on various forensic and medical diagnostic tools and methods reviewed in this article conclusively leads to the conclusion that time has come to recognize the polygraph as an admissible forensic diagnostic tool.

¹ The author is a private examiner in Israel, and a regular contributor to the publications of the American Polygraph Association. The views expressed in this column are solely those of the author, and do not necessarily represent those of the American Polygraph Association. Publishable comments and replies regarding this column can be sent to editor@polygraph.org. The APA may publish responsible comments received by the publication deadline in the following issue of the *APA Magazine*.

* Reprint from: *APA Magazine* 2013, Vol. 46(1), No. 1. The Magazine for the Polygraph Professional

Since the Frye decision in 1923 it seems that polygraph validity is haunted by what Justice Van Orsdel said: “The systolic blood pressure deception test had not gained enough standing and scientific recognition among physiological and psychological authorities to justify its admission as evidence in courts of law.”² Justice Van Orsdel’s words were repeated too many times by jurists, researchers and other interested parties sometimes disguised in a humanitarian outfit. For too many decades polygraph has been on the defense without any favorable chances. Whereas fingerprinting and DNA are evidential royalties, polygraph is not even a commoner. While being realistic enough to face the polygraph flaws, in the same token when comparing the polygraph to other forensic and medical diagnostic tools and methods recognized by courts, one cannot avoid but cry, “The King is naked.”

Latent Fingerprints

Although they are considered as “heavy weight” admissible evidence, the following case exemplifies its reliability: In the trial of *United States of America v. Byron Mitchell* (1999), a latent print examiner testified to identification between two latent prints lifted from a getaway car and the 10-print card of the defendant. The defendant claimed innocence and challenged the accuracy of the fingerprint evidence. The FBI attempted to demonstrate the scientific certainty of the identification between the defendant’s 10-print and the two latent prints found in the car. As part of the demonstration presented at trial, the FBI sent the two latent prints, together with the defendant’s 10-print, to 53 different law enforcement agencies around the United States, told them that this request was very important, and asked that their most “highly experienced” examiners determine whether any identifications could be made. This was a unique opportunity for a demonstration of concurrence among experienced examiners. Thirty-nine agencies returned analyses of the prints to the FBI. Nine of them (23%) found that either one or both of the latent prints did **not** match any of the prints from the defendant’s 10-print card.³

In 1994 the American Society of Crime Laboratory Directors (ASCLD) Proficiency Advisory Committee contacted the International Association for Identification (IAI) and asked for assistance in the manufacture and review of future testing materials. The IAI contracted with the Collaborative Testing Services (CTS), and, from 1995 to the present, the external latent fingerprint

² *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

³ L. Haber, R. Haber (2003), *Error Rates for Fingerprint Comparisons*, [in:] N.K. Ratha (ed.), *Advances in Automatic Fingerprint Recognition*, New York, Springer Verlag.

examiner proficiency test used by ASCLD has been administered by CTS, and designed, assembled, reviewed, and authorized by the IAI. Its format still consists of a number of latent prints and 10-print cards and the only responses required are identification or elimination. The summary responses reported by CTS combine consensus reports from laboratories and from individual examiners. The overall results for the seven years from 1995 to 2001 are listed in the following table:⁴

Year of Test	Number of Examiners	All Correct Responses %	One or more Erroneous ID %	One or more Missed ID %
1995	156	44	20	36
1996	184	16	3	81
1997*	204	61	6	28
1998	219	58	6	36
1999	228	62	5	33
2000	278	91	4	5
2001	296	80	3	17

* Respondent made more than one kind of error

DNA

If fingerprints are the queen of evidence then the DNA is definitely the king, or is it really? As a result of quality control case reanalysis, the Illinois State Police discovered an error rate of 25% in negative biology/DNA work. In other words, biology/DNA cases that were really positive for semen were being reported as negative.⁵ “The Houston Police Department (HPD) shut down the DNA and serology section of its crime laboratory in early 2003 after a television exposé revealed serious deficiencies in the lab’s procedures, deficiencies that were confirmed by subsequent investigations. Two men who were falsely incriminated by botched lab work have been released after subsequent DNA testing proved their innocence. In dozens of cases, DNA retests by independent laboratories have failed to confirm the conclusions of the HPD lab. The DNA lab remains closed while an outside investigation continues. In Virginia, post-conviction DNA testing in the high-profile case of Earl Washington, Jr. (who was falsely convicted of capital murder and came within hours of execu-

⁴ *Ibidem.*

⁵ Illinois State Police (2005), *DNA Testing Accountability Report*, [see in:] ww.isp.state.il.us/docs/05dnareport.pdf.

tion) contradicted DNA tests on the same samples performed earlier by the State Division of Forensic Sciences. An outside investigation concluded that the state lab had botched the analysis of the case, failing to follow proper procedures and misinterpreting its own test results.

- In 2004, an investigation by the *Seattle Post-Intelligencer* documented 23 DNA testing errors in serious criminal cases handled by the Washington State Patrol laboratory.
- In North Carolina, the *Winston-Salem Journal* recently published a series of articles documenting numerous DNA testing errors by the North Carolina State Bureau of Investigation.
- The Illinois State Police recently cancelled a contract with Bode Technology Group, one of the largest independent DNA labs in the country, expressing “outrage” over poor quality work.
- LabCorp, another large independent lab has recently been accused of botching DNA tests.

One chronic problem that is now being recognized is the uneven quality of forensic DNA laboratories. Laboratories vary greatly in the care with which they validate their methods and the rigor with which they carry them out. Another problem now emerging into the light is an unexpectedly high rate of laboratory errors involving mix-up and cross-contamination of DNA samples. Errors of this type appear to be chronic and occur even at the best DNA labs. A third problem now emerging is dishonest DNA analysts who falsify test results. I suspect this third problem is closely related to the second problem: DNA analysts are faking test results to cover up errors arising from cross-contamination of DNA samples and sample mix ups.”⁶

Of the 2,749 victims of the 9/11 WTC attack, 1,592 were identified by a variety of forensic techniques. Although the identity of the missing persons were known and although the families provided DNA comparison samples, only 111 (4%) missing persons identifications were made from the 23,608 extracted DNA samples recovered from the WTC site.⁷

⁶ J.C. Thompson (2006, January/February), *Tarnish On The ‘Gold Standard’: Recent Problems In Forensic DNA Testing*, *The Champion Magazine*, 10.

⁷ E. Lipton (2005, April 3), *At the Limits of Science*, *New York Times*.

Comparative Bullet Lead Analysis (CBLA) a.k.a Compositional Analysis of Bullet Lead (CABL)

The CBLA was first used in 1963 in JFK's assassination investigation. The CBLA matches the chemical composition of a bullet lead found in the scene with bullets and/or bullet box found in the suspect's possession, under the assumption that the molten (melted lead alloy) source has a uniform chemical composition throughout, so no two molten sources have the same chemical composition. CBLA is considered a scientific, flawless and accurate technique. Since its first use the FBI Crime Lab performed about 2,500 analyses that led to conviction. Due to on-going critique by defense lawyers and the press the FBI finally asked the United States National Academy of Sciences to research the scientific merit of the process. The Academy conclusion was that, "Variations among and within lead bullet manufacturers make any modeling of the general manufacturing process unreliable and potentially misleading in CABL comparisons."⁸

Medicine

Research regarding the cause of diagnostic error in medicine found that, "We argue that physicians in general under-appreciate the likelihood that their diagnoses are wrong and that this tendency to overconfidence is related to both intrinsic and systemically reinforced factors."⁹ The error rate in some instances is alarming.¹⁰

Some excerpts from the research:¹¹

- "126 patients who died in the ICU and underwent autopsy, physicians were asked to provide the clinical diagnosis and also their level of uncertainty. Clinicians who were 'completely certain' of the diagnosis before death were wrong 40 percent of the time."

⁸ United States National Academy of Sciences' Board on Chemical Sciences and Technology, (2004), *Forensic Analysis Weighing Bullet Lead Evidence*, The National Academies Press, Washington D.C.

⁹ E.S. Berner, M.L. Graber (2008), *Overconfidence as a Cause of Diagnostic Error in Medicine*, *The American Journal of Medicine*, (121)5, S2.

¹⁰ *Ibidem*.

¹¹ *Ibidem*.

Conditions	Findings
Pulmonary TB	Studies that have specifically focused on the diagnosis of pulmonary TB; 50% of these diagnoses were not suspected ante-mortem
Pulmonary embolism	Of 67 patients who died of pulmonary embolism, the diagnosis was not suspected clinically in 37 (55%)
Ruptured aortic aneurysm	Of 23 cases involving abdominal aneurysms, diagnosis of ruptured aneurysm was initially missed in 14 (61%); in patients presenting with chest pain, diagnosis of dissecting aneurysm of the proximal aorta was missed in 35% of cases
Subarachnoid hemorrhage	Updated review of published studies on subarachnoid hemorrhage: 30% are misdiagnosed on initial evaluation
Cancer detection	Of the 250 malignant neoplasms found at autopsy, 111 were either misdiagnosed or undiagnosed
Breast cancer	50 accredited centers agreed to review mammograms of 79 women, 45 of whom had breast cancer; the cancer would have been missed in 21%
Melanoma	Second review of 5,136 biopsy samples; diagnosis changed in 11%
Bipolar disorder	The initial diagnosis was wrong in 69% of patients with bipolar disorder and delays in establishing the correct diagnosis were common
Appendicitis	Retrospective study at 12 hospitals of patients with abdominal pain and operations for appendicitis. Of 1,026 patients who had surgery, there was no appendicitis in 110 (10.5%); of 916 patients with a final diagnosis of appendicitis, the diagnosis was missed or wrong in 170 (18.6%)
Cancer pathology	The error rate of pathologic diagnosis was 2%–9% for gynecology cases and 5%–12% for non-gynecology cases
Endometriosis	Digital videotapes of laparoscopies were shown to 108 gynecologic surgeons; the inter-observer agreement regarding the number of lesions was low (18%)
Psoriatic arthritis	1 of 2 SPs with psoriatic arthritis visited 23 rheumatologists; the diagnosis was missed or wrong in 9 visits (39%)
Atrial fibrillation	Review of automated ECG interpretations read as showing atrial fibrillation; 35% of the patients were misdiagnosed by the machine, and the error was detected by the reviewing clinician only 76% of the time
Infant botulism	Study of 129 infants in California suspected of having botulism during a 5-yr period; only 50% of the cases were suspected at the time of admission

- Lack of knowledge *per se*, such as seeing a patient with a disease that the physician has never encountered before. More commonly, cognitive errors reflect problems gathering data, such as failing to elicit complete and ac-

curate information from the patient; failure to recognize the significance of data, such as misinterpreting test results; or most commonly, failure to synthesize or put it all together.

- “The breakdown in clinical reasoning often occurs because the physician isn’t willing or able to ‘reflect on [his] own thinking processes and critically examine [his] assumptions, beliefs, and conclusions.’ In a word, the physician is too ‘confident.’”

Courts

Gelman et al. (2004) examined 4,578 appeals of death sentences in U.S. states between 1973 and 1995 and found that, “... the overall rate of prejudicial error in the American capital punishment system was 68%. In other words, courts found serious, reversible error in nearly 7 of every 10 of the thousands of capital sentences that were fully reviewed during the period... Capital trials produce so many mistakes that it takes three judicial inspections to catch them leaving grave doubt whether we do catch them all. After state courts threw out 47% of death sentences due to serious flaws, a later federal review found ‘serious error’ undermining the reliability of the outcome in 40% of the remaining sentences.”¹²

Psychiatry

In order to avoid punishment in criminal trials a defendant’s use of the insanity defense is where s/he claims not being responsible for her/his actions due to mental health problems which are being determined by psychiatrists. The “Rosenhan experiment”¹³ examined the validity of psychiatric diagnosis. The study consisted of two parts. The first involved eight “pseudo-patients” – people who had never had symptoms of mental disorder – who, as part of the study, briefly reported auditory hallucinations in order to gain admission to psychiatric hospitals across the United States. After admission, the pseudo-patients no longer reported hallucinations and behaved as they normally would. The pseudopatients remained in hospital for 7 to 52 days. None of the pseudo-patients were detected, and all but one were admitted with a diagnosis

¹² A. Gelman, J.S. Liebman, V. West, A. Kiss (2004), *A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States*, *Journal of Empirical Legal Studies*, 1(2), 209–261.

¹³ D.L. Rosenhan (1973), *On Being Sane in Insane Places*, *Science*, Vol. 179, 250–258.

of schizophrenia and were eventually discharged with a diagnosis of schizophrenia in remission. Although they were not detected by the staff, many of the other patients suspected their sanity (35 out of the 118 patients voiced their suspicions). In the second part of the experiment staff at a teaching hospital, who had learned of Rosenhan's above results, were informed that one or more pseudo-patients would attempt to be admitted to their hospital over an ensuing three-month period. Out of the 193 admitted 41 patients were subsequently identified as likely pseudo-patients but in fact no pseudopatient had been sent at all.

Polygraph Compared to Other Forensic and Diagnostic Tools

Widacki & Horvath (1978) examined in laboratory conditions the relative validity of the polygraph with three other common methods of criminal investigation. The table below represents their findings.¹⁴ Crewson (2003)¹⁵ reviewed 1,158 articles and abstracts (145 fit the objectives of the literature review, yielding data on 198 studies) which compared the validity of polygraph with other medical and psychological screening and diagnostic tools. The comparison revealed the following data:

Diagnostic Tool	Correct	Incorrect	Inconclusive	False Positive
Polygraph	90%	5%	5%	1.30%
Handwriting	85%	5%	10%	1.40%
Eyewitness	35%	20%	45%	9.10%
Fingerprint	20%	0%	80%	0%

Diagnostic Tool	Sensitivity	Specificity	Combined	Studies (N)
Polygraph (Diagnostic)	0.92	0.83	0.88	37
MRI	0.86	0.88	0.87	17
CT	0.83	0.89	0.86	19
US	0.84	0.87	0.86	38
X-Ray	0.77	0.85	0.81	12
MAST	0.64	0.92	0.78	3

¹⁴ J. Widacki, F. Horvath (1978), *An Experimental Investigation of the Relative Validity and Utility of the Polygraph Technique and Three Other Common Methods of Criminal Investigation*, *Polygraph*, 7(3), 215–222.

¹⁵ P.E. Crewson (2003), *Comparative Analysis of Polygraph with other Screening and Diagnostic Tools*, *Polygraph*, 32 (1), 57–85.

Polygraph (Screening)	0.59	0.9	0.74	2
DSM-IV	0.72	0.68	0.7	1
MMPI	0.68	0.65	0.67	17
MMPI (Screening)	0.7	0.53	0.61	5

see next page for a description of diagnostic tools

In addition, a reliability (inter-rater agreement) comparison was made and revealed the following data:

	Polygraph	Medicine	Psychology
Agreement	91%	81%	88%
No. Subjects	102	150	174

Validity and reliability of the polygraph

A compendium of various researches done by Ansley in 1983 and later in 1990 averages the validity and the reliability of the polygraph around 94%.¹⁶

Sensitivity – The proportion of diseased cases with a positive test (perfect accuracy = 1.0) i.e. the test identifies the sick. Polygraph identifies guilty subject as guilty.

Specificity – The proportion of non-diseased cases with a negative test (perfect accuracy = 1.0) i.e. the test identifies the healthy. Polygraph identifies innocent subject as innocent.

MRI – Magnetic resonance imaging is a test that uses a magnetic field and pulses of radio wave energy to make pictures of organs and structures inside the body.

CT – Computerized tomography imaging, also referred to as a computed axial tomography (CAT) scan, involves the use of rotating x-ray equipment, combined with a digital computer, to obtain images of the body.

US – Ultrasound imaging is a common diagnostic medical procedure that uses high-frequency sound waves to produce dynamic images (sonograms) of organs, tissues, or blood flow inside the body.

¹⁶ N. Ansley (1983), *A Compendium on Polygraph Validity*, *Polygraph*, 12(2), 53–61; N. Ansley (1990), *The Validity and Reliability of Polygraph Decisions in Real Cases*, *Polygraph*, 19(3), 169–181.

MAST – Michigan Alcohol Screening Test is one of the oldest and most accurate alcohol screening tests available.

MMPI – The Minnesota Multiphasic Personality Inventory is one of the most frequently used personality tests in mental health. The test is used to assist in identifying personality structure and Psychopathology.

DSM IV – Diagnostic & Statistical Manual of Mental Disorders, 4th Edition, is a manual published by the American Psychiatric Association (APA) that includes all currently recognized mental health disorders.

Additional Pro-Polygraph Considerations

The efficiency of the polygraph as an investigative tool aid was demonstrated in the Light and Schwartz (1993) study. 1,069 forensic examinations involving 920 felony investigations conducted in the second half of 1990 by the US Army CID were surveyed. Eight primary forensic disciplines that were used in support of the investigations were used in this study. Of those forensic examinations, 584 (55%) were in traditional laboratory disciplines combined, and 485 (45%) were with the polygraph. The polygraph provided the investigator with 432 (89%) opinions that contained positive results and the laboratory disciplines provided positive results in 431 (74%) examinations.

Tool	N	Positive Results	Negative Results
Polygraph	485	89%	11%
Latent Finger Prints	154	59%	41%
Questioned Documents	145	72%	28%
Illicit drugs	133	93%	7%
Firearms	51	76%	24%
Trace Evidence	51	65%	35%
Serology	40	85%	15%
Photographic	10	50%	50%

The polygraph was the most utilized and effective of the individual disciplines.¹⁷

¹⁷ G.D. Light, J.R. Schwartz (1993), *The Relative Utility of the Forensic Disciplines*, Department of Defense Polygraph Institute, Fort McClellan AL.

Additional considerations:

- Due to its mobility polygraph tests can be executed almost everywhere without the need of an expensive laboratory.
- In comparison to fingerprints and DNA, which disqualifies over 50% of specimens, almost none are disqualified by the polygraph (inconclusive tests are solved by re-examinations). Yet, if examiners want to play it safe and increase the numerical scoring threshold the error rate would be around 2%.¹⁸
- The fact that about 69% of specific test examinees are found truthful¹⁹ leads to the conclusion that the polygraph assists the innocent to prove their innocence.

Final Note

The data detailed in this article bear no intention of discrediting any of the described diagnostic tools or methods. It is offered to highlight the fact that tools and methods which are less accurate than polygraph are being accepted by the legal community while the polygraph is not, which in turn raises the question: why? Some apparent explanations might be:

- Conservatism – Insistence on preserving the legal framework results in conservatism (and in some rare instance in stagnation), and to a slower pace of adopting innovation (take the Frye precedent as an example). “Beyond a shadow of a doubt” exemplifies the legal system point of view.
- Fear of Unemployment – No matter how simplistic it sounds the fact remains that excessive utilization of polygraph tests to determine guilt will downsize the legal system workforce as automation did to production lines.
- Self – Preservation – Nowadays the legal system is more concerned with protecting and preserving its bureaucratic procedures than making justice as in setting free a serial rapist or a serial murderer for technicalities and letting him endanger society. The objective of making justice was sacrificed in favor of sacred means.
- Unrealistic Expectations – Unlike courts that decide upon an accused’s guilt after tedious long sessions, polygraph examiners do it in about an

¹⁸ D.J. Krapohl (1998), *A Comparison of 3 and 7 Point Scale, Polygraph*, (27)3, 210–218; E.M. Harwell (2000), *A comparison of 3 and 7 Position Scales, Polygraph*, (29)2, 195–197.

¹⁹ E. Elaad, M. Kleiner (1992), *The Police Use of Polygraph Examinations in Israel, Policija i Sigurnost* (Police and Security), (1)6-5, 418–430.

hour. This speedy decision making on such sensitive issues creates normally a very high level of expectation to the point of requiring 100% accuracy, which in return leads to zero tolerance of mistakes, but a 100% accuracy, the polygraph cannot provide.

While these explanations are but circumstantial, the direct and last nail in the polygraph's inadmissibility coffin is:

- Exclusivity (Monopolism) – Unlike the polygraph examiner, all forensic experts' opinion points **indirectly** at the accused's guilt. For example a specimen of an accused latent fingerprint and/or DNA found in the scene of crime does not prove her/his guilt but merely her/his presence there, providing an undiscriminating logical explanation to the accused's presence obsoletes the forensic expert's opinion.

The polygraph examiner expert's opinion is the **only** forensic expert opinion that actually points **directly** at the accused's guilt. By doing so the examiner penetrates the court authority and interferes in a decision that is exclusively granted to the court and juries.

Conclusion

"Imperfection is the only perfection"

Polygraph should not be judged in absolute terms but in relative terms and its relative accuracy is at least as good if not better than any other forensic diagnostic and non-diagnostic tools or methods which are being accepted as admissible evidence.

Approximately two out of three examinees are found truthful which from a social point of view is significant, especially when an accused has no other mean to prove her/his innocence.

Polygraph is the most cost-effective (time, cost, results, availability) diagnostic tool. In an era where sacred cows, admissible evidence, is bleeding (some of them to death), the time has come for the legal community to embrace the polygraph and use it in its quest for truth, internalizing Aristotle's perception: **"The whole is greater than the sum of its parts."**