EUROPEAN POLYGRAPH

PUBLISHED QUARTERLY

Volume 4 | 2010 | Number 3 (13)

Andrzej Frycz Modrzewski Krakow University

EUROPEAN
POLYGRAPH
Volume 4 • 2010 • Number 3 (13)

Contents

Book reviews

EUROPEAN
POLYGRAPH

Volume 4 • 2010 • Number 3 (13)

From the Editor

Traditionally, we published four volumes of our magazine a year. Two single volumes and a double one. This year, for the first time, we are publishing four single volumes, with No. 3 including, besides reviews, only one study, yet one greatly exceeding the volume we initially defined for articles, which as you know is 12 pages. This is a study entitled *Polygraph Examination as Scientific Evidence* by Jerzy Konieczny.

In future, we would like to continue publishing four individual volumes a year, with each year's No. 3 allowing the publication of more spacious monographic studies which would fill up the majority of the space devoted to the articles and not be limited to 12 pages.

The reason for this is that the Editing Board recognises that a valuable work of a larger volume, whose subject deserves publication in a magazine devoted to polygraph studies, may crop up at least once a year. Providing the option of publishing such longer works in the field is also an incentive for future authors to write not only brief articles, but also broader theoretical and analytical works. There have so far been problems with publishing such spacious works in academic or specialist magazines.

By beginning the new tradition with the excellent work by Jerzy Konieczny, we are at the same time trying to establish the expected level of studies we want to publish in the future.

Jan Widacki
Editor-in-Chief

European Polygraph
Volume 4 • 2010 • Number 3 (13)

Jerzy Konieczny*
Andrzej Frycz Modrzewski
Krakow University
Krakow, POLAND

Polygraph Examination as Scientific Evidence

I

I would like to begin my examination of the problem at hand by focusing on the meaning of the two terms used there, namely *polygraph examination* (further referred to as PE) and *scientific evidence*.

I understand PE to be an action performed by an expert, the outcome of which is a possible indication of deception of the subject of the examination with regard to a certain topic of relevance under the law. I will only discuss PE that is undertaken as a part of a certain legal procedure and for the purposes of this procedure; therefore, the outcome of the PE will have the status of evidence in the said procedure.

It is not easy to define scientific evidence. I will, for now, assume that scientific evidence belongs to the category of expert evidence, in the sense that scientific evidence constitutes a portion of what is considered expert evidence. I am go-

* jerkonieczny@wp.pl

ing to argue that scientific evidence differs from other types of expert evidence in how its claims are formulated, interpreted and justified, as well as in the necessity to give regard to certain rules concerning meta-evidence.

An analysis of the relationship between PE and scientific evidence should address the following questions: (1) What are the methodological characteristics of PE? (2) What is the nature, from the point of view of methodology, of the knowledge applied in PE? (3) Is PE scientific evidence, and if it is, then how so? The aim of this paper is to attempt to answer these questions. In the attempt, I intend to rely strongly on the general body of knowledge of forensic science, even where it fails to apply directly to PE.

I also assume that the knowledge which is used by the expert performing PE may be divided into two categories: (a) practical knowledge, acquired from those who trained the expert in the profession, from quality controllers, from colleagues, as well as professional experience, amassed hands-on in the course of the career in the field of PE; and (b) theoretical knowledge, acquired mainly from professional publications. The categories are hardly exclusive. The division, although more of a typology, holds certain usefulness for the analyses presented herein.

II

I will assume that PE-related issues belong to forensic science, and that forensic science is empirical and scientific.

The central category of my analysis here is *science*, understood as a category of knowledge, i.e. a product of enquiry. It is of crucial importance to distinguish between the knowledge that constitutes science and the knowledge that does not. I will abstain from summarising the entire fundamental dispute of the philosophy of science, namely that of the demarcation of science and non-science. I will only point to one aspect of this issue.

The issue of the criteria of said demarcation arose in the dispute of the two great methodological approaches, the positivist approach and the hypothesis-oriented approach, Karl Popper style. The dispute centred on deciding which constitutes a better criterion: the ability to verify or the ability to falsify a system of claims of an empirical science. The need to test such a system empirically however remained unchallenged with regard to claims that were supposed to constitute hypotheses or scientific truths. Both positivism and the hypothesis-oriented approach fully agree also with respect to the fundamental significance of testing hypotheses at the empirical level.

I thus follow a well-founded tradition here when I assume that empirical testability is a fundamental characteristic of products of empirical sciences, and that testability here refers specifically to inter-subjective testability.

Since claims exist through language, it is crucial to ensure that selected terms are understood uniformly by competent examiners (i.e. to ensure inter-subjective communicability and meaningfulness). Formulation of a scientific question and the description of its solution should be executed in a manner that is inter-subjectively meaningful/communicable. Then it will also be possible to ensure inter-subjective testability of claims: the testability will be available to anyone with generic competent intelligence, where the competence is acquired through appropriate training, and defined by a description of that training. Essentially then, a claim that is proposed to be scientific should be constructed in such a manner that it is subject to inter-subjective testing, i.e. testing that should be possible to perform for any researcher in a given area of science, that should follow the same course in every case, and that should, barring possible errors, lead to the same conclusions in every case.

It is also evident to an observer of research practice that the criterion of inter-subjective meaningfulness and testability, while generally useful, is not extremely precise. In particular, it allows for the option of deciding that certain claims, and therefore certain theories, may be scientific to a certain degree. The demarcation is therefore not limited to the dichotomy of science versus non-science. It allows for a gradation of the quality of being scientific, i.e. for an assessment that one claim is "more scientific" than another.

No argument is being made here that inter-subjective communicability and testability are the only criteria for a claim being scientific. However, an attempt to determine other criteria, or even just an attempt to address the relevant notions already present in the academic discourse, would necessitate sailing the endless ocean of philosophy of science and losing the focus on the subject matter of this paper. Suffice it to assume that scientific knowledge exists and that a part of it concerns PE.

I would like to present one more position here, with regard to the claim of the expert, and specifically to a claim in a specific case that the expert has arrived at in a manner befitting the criteria of inter-subjective meaningfulness and testability, and that is going to be used as evidence or is already being used as such. Does this claim belong to the realm of science, i.e. does it constitute a part of forensic science? The answer is negative, for two reasons. Firstly, forensic science is a nomothetic science, one which formulates general claims and descriptive generalisations as well as optimising, explanatory and other generalisations. Expert evidence on the other hand is always generated for the purposes of a single case, which may or may not be criminal in nature, as a result of an examination of a single fact, and as such cannot be classified as science. It may seem strange to claim that an expert working on a specific case always focuses on a single fact, where evidently many complex research endeavours (including PE) require for their completion numerous and complicated research activities. Nonetheless, they always pertain to one issue. Should it be necessary, it may be agreed that the assumption of the single issue being at the focal point is an idealisation, and only make it specific as the need arises. The matter however exceeds the scope of this paper. An analogy with medicine seems more useful: examination of a single patient, even if it is complex and far-reaching, is not a scientific act on the part of the medical professional, and the result of the examination does not belong to the science of medicine, understood as a system of generalised claims. Naturally, a single casus may constitute and often does constitute inspiration for research that is instructive and even break-through in nature (e.g. in the cases of Frey, Daubert, etc.). Nonetheless, no specific expert evidence constitutes science, even though it is sometimes described as scientific. It is one of the key goals of this paper to determine under which circumstances this description is accurate with regard to PE.

I would like to note two more points under the heading of introductory remarks. Firstly, I will not discuss in this paper the legal issues surrounding admissibility of PE. I will only indicate selected specifics thereof when methodologically relevant. Secondly, I will assume that the reader is familiar with PE at least at a fundamental level.

III

Forensic evidence may not be introduced into a legal procedure other than through an expert (Kiely 2001: 44), and therefore evidence may only be present there as delivered by a person with specialized knowledge. This knowledge consists of certain skills, of training in how to apply these skills, and of professional experience (Freckelton 2000: 713). Under such circumstances expert evidence enters into play, usually in the form of an expert opinion. All of the above applies to PE as well.

Experts can be classified in numerous ways, (see e.g. Speight 2009: 6 onward). For the purposes of this paper, the establishment of two categories is particularly useful, namely those of an expert consultant and an expert witness. The expert consultant provides consultation during the investigation; assists the prosecutor in developing the trial tactics; provides clarification to the judge with regard to the contents of the expert opinion, etc. The expert consultant is not formally involved in the proceedings, and the presence of such an expert may not be publically acknowledged; various laws regulate this matter differently. The matter looks different with an expert witness. The expert witness files a formal testimony with the court, is obliged to give an account of all the material used, may be questioned by the parties, etc. (Matson, Daou and Soper 2003: 8–10). PE experts may work in both capacities, but the distinction is without consequence for the methodological characterisation of PE.

According to another definition, "an expert is a witness who possesses those qualifications that permit him to offer to the court not only observational information but also to formulate opinions and draw conclusions on the basis of an examination of forensic data using knowledge 'beyond that of the average juror'" (Moenssens 2009: 1012).

Moenssens definition includes one very important element. Namely, it points to the possibility of using the expert's qualifications in a twofold manner. The expert may not only present opinions on facts, but also interpret them, in the sense that the expert may draw conclusions independently. Testimony of a witness who is not an expert is limited to relating what the witness did, heard or saw, and no drawing of conclusions on this basis is allowed. Even if a testimony contains some interpretative statements (in the sense referenced above), such interpretation may not exceed general knowledge (Freckelton 2000: 716). Comments are sometimes made to the effect that it is actually impossible to fully separate fact from interpretation, if only because the witness's body language is accessible to the court (Jones 1994: 102–103), but this should have no bearing on the present analysis. An expert polygraphist, as will shortly be recounted, always presents observational information together with formulating opinions, where the opinions constitute conclusions from the examination previously undertaken by said expert.

The distinction between fact evidence and opinion evidence, which used to hold great importance from the legal standpoint, has continued to be analysed, but the importance assigned thereto has declined recently. Choo explains how, regardless of the theoretical approach, the matter is of practical significance (2006: 251). Roberts and Zuckerman, who oppose a dogmatic approach to the issue, suggest that the evidentiary value of facts and opinions should be held paramount, and the decision as to what constitutes facts and what constitutes opinions should be left to the court (2004: 146). Dwyer (2008: 74–76) claims that contemporary methods of reasoning deprive this distinction of its epistemological basis and that insisting on it is an operational convenience rather than a valid distinction. He also however emphasises that expert evidence is "opinion" rather than "fact", and that this sets it apart from other legally admissible options. This is hard to disagree with. However, it seems just as correct to observe that very often the result of the expert's work is a hybrid of observation and conclusions on the basis thereof, i.e. opinions. "Finally, the archetypal expert – not a percipient witness, retained by a party, and lacking prior case-specific knowledge – also often offers 'hybrid' testimony in the sense that he may provide both factual information and opinion" (Kaye, Bernstein, Mnookin 2004: §3.2.2).

I assume that PE may be performed to achieve either of the following two goals: (a) to decide whether the subject is giving honest answers to the critical questions in the tests, and (b) to decide whether the subject recognizes certain events.

The PE expert performing the examination undertakes numerous preparatory activities: familiarises himself/herself with the case in the course of which he/she is to carry out the examination; conducts pre-test interviews and goes through chart collections leading to obtaining polygrams; performs and assesses the polygrams; finally, draws conclusions. Polygrams are a record of certain information, assessment is a reading of this information, and the drawing of conclusions constitutes an interpretation of this assessment. Therefore: charts collection + assessment of polygrams = establishment of facts, and drawing conclusions from the facts thus established = formulation of an opinion. The above two "equations" are crucial for a methodological characterisation of PE. I would like to concentrate now on discussing them in more detail.

A PE report must include, *inter alia*, a presentation of the facts established, and the conclusions drawn, i.e. it must include "both factual information and opinion"; a polygraphist undoubtedly is the "archetypal expert".

The division between establishing facts and formulating opinions used to be referred to as the division between the learning sphere and the decision-making sphere in performing the tasks of an expert.

Further considerations necessitate first that the notion of *fact* be made more precise. Interestingly, theoreticians of evidence pay little attention to defining *fact* precisely, making do with, for instance, the fairly general statement that facts pertain to "observable qualities of the world" (Ho 2008: 7). Dwyer quotes the dictionary definition of what a fact is: "a particular truth known by actual observation or authentic testimony, as opposed to what is merely inferred, or to conjecture or fiction; a datum of experience, as distinguished from the conclusions that may be based upon it". He adds that a fact "becomes a thing associated with certainty of knowledge" (Dwyer 2008: 87). There is also the approach that "facts alone, even scientific facts, are not knowledge". They only become knowledge after the following questions are answered: "What range of facts is worth investigation? What is the proper way to investigate them?

What do the results of the investigation mean?" (Beecher-Monas 2007: 50). The above comments further support the view that the immediate results of an expert's work on a specific case do not constitute science.
For the moment, let us remain with the elementary intuition that a fact is a situation where a certain object (x), where the variable (x) may apply to a group of persons, has a quality P to a non-zero degree, in short: P(x). Let us call this statement F. Statement F may be completed by the addition of data on place p and data on time t, referring to the occurrence of the fact. Should this be necessary, the notation is: $P(x)_{p,t}$. In a specific PE, the expert may decide, for example, that a specific person a_1 (let us call this person John Doe) has the quality P.

The statement
F: A certain object (a_1) has the quality P to a non-zero degree
has the following qualities: (1) syntactically, it is a sentence, and its content is a proposition, (2) in order for this statement to possibly carry true information, it must be equipped in meaning, and thus become understandable; if I am allowed to be somewhat caustic here, it should be a *meaning*, singular, rather than *meanings*, plural; in any case, the need for an expert's statement to be unambiguous bears emphasising; (3) the fact that a statement is understandable has no bearing on whether it is true or false, (4) the statement announces that a certain individual (a_1) belongs to a certain set, in this case: to a set of object with the quality P, and at this time it is irrelevant how P is measurable, (5) contains one individual constant and one one-place predicate, and (6) contains neither quantifiers nor conjunctions [and, or, if].

Let us consider the following types of situations:

(i)
An expert conducted an examination of a person (a_1) using one of the comparison question techniques (CQT) and performing an appropriate number (3 or 5) of Utah PLT tests. Then, by means of an analysis of the charts collected and by using the 7-point Backster scale he/she calculated that the grand total of the examination is minus 12, which led the expert to formulate the obvious observation that: Person (a_1) belongs to the set of persons who obtain, in an examination conducted with the use of PLT tests, a grand total of minus 12 points. Of course the actual result of the examination might be different, placing the person at minus 2 or plus 15. Therefore, we can generalise and say that the observation of an expert in this case, i.e. where the examination is using

CQT, has the following form: Person (a_1) belongs to the set of persons who obtain, in an examination conducted with the use of Utah PLT tests, a grand total of n points. We thus arrive at the following F-type statement:
F1: Person (a_1) belongs to the set of persons who obtain, in an examination conducted with the use of Utah PLT tests, a grand total of n points.

(ii)
An expert conducted an examination of a person (a_2) using the CIT technique and performing an appropriate number (no fewer than 4) of POT tests. Then, by means of an analysis of the charts collected and by using the Lykken method he/she calculated that the grand total of the examination is plus 6. Of course the actual result of the examination might be different, placing the person at, let us say, 4. Therefore, we can generalise and say that the observation of an expert in this case, i.e. where the examination is using CIT, has the following form: Person (a_2) belongs to the set of persons who obtain, in an examination conducted with the use of CIT tests, a Lykken number of m points. We thus arrive at another F-type statement:
F2: Person (a_2) belongs to the set of persons who obtain, in an examination conducted with the use of CIT tests, a Lykken number of m points.

{**DG 1.** I am using the example of Utah PLT and POT tests without explaining the details of their application. A reader who is a polygraphist will easily understate the examples. There is no need to introduce examples of other techniques and methods; the results would be the same.[1]}
We thus have the following examples of F-type statements:
F1: Person (a_1) belongs to the set of persons who obtain, in an examination conducted with the use of Utah PLT tests, a grand total of n points.
F2: Person (a_2) belongs to the set of persons who obtain, in an examination conducted with the use of CIT tests, a Lykken number of m points.

In the context of PE it is difficult to agree with the comment of Dwyer cited above, to the effect that a fact "becomes a thing associated with certainty of knowledge". There is no certainty here as to the values of m and n. Nonetheless, the content of the above statements is a proposition that a certain fact has occurred. These statements satisfy the criteria listed in (1)-(6) above. Such

[1] The author believes, as do many others, that footnotes make reading difficult. However, there are comments that, while of secondary importance to the main reasoning, should nonetheless be made. Such comments will be placed in the digressions, denoted by the DG symbol and numbered, and delimited by the symbols {...}.

sentences, and in particular sentences that satisfy conditions (4), (5) and (6), are atomic sentences (Ziembiński). Moreover, since they relate to a singular specific observation (and we noted earlier that PE, like any expert research activity, is focused on a single fact), we may say that they are observational statements. I will refrain here from addressing the issue of the existence of so-called pure facts. We will also not engage, at least for now, in consideration of the adequacy of interpretation of these statements. It appears that an introspective notion of interpretation, of the type: "right here and right now the expert believes the following", should be best. Please note also that statements F1 and F2 could easily be appended with notes relating to the time and place of the facts under description. This would lead to sentences that are basic, as understood by Karl Popper. We therefore obtain the following:

When an expert produces a basic atomic sentence, he/she makes a claim referring to a fact.

Let us observe at this point that none of the statements F1-F2 achieves the goal of PE, i.e. gives answers to the questions whether the subject is giving honest answers to the critical questions in the tests, and/or whether the subject recognizes certain events. In order to achieve these goals, a different knowledge is needed than in order to establish facts. This knowledge is a set of principles of formulating opinions. These principles are connected with examination techniques and in the case of the above examples F1-F2 read as follows:

(i)

With regard to Utah PLT, the principles of formulating opinions are as follows. Let us assume for the sake of simplicity that reactions that occurred in the course of examination of person (a_1) were, for the relevant questions, always stronger than for comparison questions, or that they were, for the relevant questions, always less strong than for comparison questions. In that situation, the principle is: "[T]he scores are then summed to provide a total score for the test, and the outcome is based on this total. If the total is -6 or lower, the outcome is deceptive, if the total is +6 or higher, the outcome is truthful; totals between -5 and +5 indicate an inconclusive outcome" (Raskin, Honts 2002: 19–20). In other words: (a) if the number of Backster points is $m \leq -6$, formulate the opinion DI; (b) if $m \geq +6$ formulate the opinion NDI; (c) if $m \leq +5$ and at the same tie $m \geq -5$ formulate the opinion IC. We thus have the following option of opinion O for F1:

F1: Person (a_1) belongs to the set of persons who obtain, in an examination conducted with the use of Utah PLT tests, a grand total of n points.

If we assume n = -12 then:

O1a: Person (a_1) gave dishonest answers to relevant questions of the tests.
If we assume n = +8 then:
O1b: Person (a_1) gave honest answers to relevant questions of the tests.
If we assume n = -2 then:
O1c: Test of person (a_1) remains inconclusive.

{**DG2.** Issues of principles of formulating opinions based on computer algorithms and probability analysis will be discussed in the next chapter.}

(ii)
With regard to the technique of CIT, which consists in conducting a series of POT tests, the expert in each test determines a question, other than the first two questions, which caused the strongest reaction. If this is a reaction to a key question, the test scores 2 points. If it is the second strongest reaction, the test scores 1 point. This is how the Lykken number is calculated. If the total number of points in the whole examination exceeds by at least 1 the number of tests conducted (n_{POT}), i.e. if n is greater than (n_{POT}), then it is assumed that the subject recognizes the event. If not, it is assumed that the subject does not recognize the event. Some researchers dilute this formula and allow, in place of "if n is greater than (n_{POT})", "if n is greater than (n_{POT}) or equal to (n_{POT})". I will not discuss this matter further here. A reader may find further comments on it in Lykken 1981 and other works, such as for instance Krapohl, McCloughan, Senter 2006: 127). We thus have the following:

O2a: The subject (a_2) recognizes a (certain) event, or
O2b: The subject (a_2) does not recognize a (certain) event.

The following statements are then examples of opinions:
O1a: Person (a_1) gave dishonest answers to relevant questions of the tests.
O1b: Person (a_1) gave honest answers to relevant questions of the tests.
O1c: Test of person (a_1) remains inconclusive.
O2a: The subject (a_2) recognizes a (certain) event.
O2b: The subject (a_2) does not recognize a (certain) event.

{**DG 3**. In the case of polygraph screening examination, where multiple issues formats should be applied, the following opinions are also allowed: "No Significant Reactions/Responses (NSR)", "Significant Reactions/Responses (SR)" and "No Opinion (NO)" or "Inconclusive (INC)". However, if the examinations result in the occurrence of "significant responses", a specific issue test should be conducted. Legislation allowing, "professional opinion that an *examinee* *was deceptive*, based on physiological data, should only result from a specific issue test." (Model Policy 2010). The above comments are without effect on the sense of the reasoning presented. However, doubt remains as to whether solely revealing the occurrence of Significant Reactions/Responses (SR) is an opinion or just a statement of fact. These issues will further be addressed in the following chapters.}

If an expert produces a basic atomic statement, he/she says something with regard to a fact. Basic statements are observational. Observational basic sentences may be used to support other statements, non-observational ones, which are then called empirical. We will not go into the details here of how empirical sentences are justified. Let us however just note the following:

When an expert produces an atomic non-observational empirical statement, he/she presents an opinion.

Taken jointly, the above comments constitute a proposition on how to solve the problem of the fact/opinion distinction in PE.

The knowledge necessary to formulate an opinion is of course significantly different from the knowledge necessary to establish facts. This allows us, or even prompts us, to distinguish between two spheres: the learning sphere and the decision-making sphere.

{**DG 4.** The author was inspired to discuss the distinction between the learning and decision-making sphere with regard to expert evidence when reading the works of Professor Kazimierz Jaegermann (1921–1988), who was an outstanding Polish forensic medicine specialist and theoretician of expert knowledge. Unfortunately, he never published in English. Jaegermann believed that taking measurements and describing them, i.e. establishing facts pertaining to a specific portion of reality, was characteristic of the former stage. The latter stage, i.e. the decision-making stage, consisted in the interpretation of data gained in the first stage, by means of relating the data to knowledge, and specifically principles of formulation of opinions. The said principles outline the so-called levels of aspiration, i.e. criteria allowing a certain type of formulation of opinion, useful for the recipient of PE, namely the lawyer (Jaegermann 1991).

In the case of PE, the portion of reality would be delimited by the goals of the specific examination, which would consist of a pre-test interview but also, most importantly, of charts collection. As for the principle of formulation of opinions, an example could be the number cited in the examples referenced above, such as for instance the scoring using the Backster scale. Depending on the outcome of the learning stage, an expert will either arrive at the level of aspiration to issue a DI or NDI opinion, or will not arrive at any of these levels and will have to present the examination as IC.}

Let us briefly address one more matter. Can statements like the F- and O-statements cited above be inter-subjectively communicable (meaningful) and inter-subjectively testable? A statement is inter-subjectively testable if all experts understand it in the same manner and if testing it is available to practically all and any representative of the given science, proceeds analogically, and – barring any errors – leads to the same outcome.

It seems that ensuring inter-subjective communicability of F- and O-type statements should, at least theoretically, be easy. It is clearly a matter of language, of the terminology used, of how terms and notions are defined, and of the principles of inference, including principles of formulating opinions. Many efforts have been undertaken by polygraphists in this area. They include standardisation of examinations, certification of experts, accreditation of laboratories, not to mention publications aimed at regulating the language of PE. The first fundamental publication in this realm, updated later, was published in 1997 (Krapohl, Sturm 1997). Currently, besides other current literature, the matter is regulated by the standard ASTM E 2035 (Terminology Related to Forensic Psychophysiology). This of course in no way means that all polygraphists with no exception follow the suggestions made therein. Unfortunately, lapses in terminology use do occur. However, they are usually clearly recognizable and thus easy to correct and amend. Let us assume then that:

It is possible to conduct PE and present a report from the examination in a manner that is inter-subjectively communicable.

The problem of inter-subjective controllability is much more difficult. This is due to the fact that PE is an act that cannot be repeated, in the sense that the very performance of the examination "changes the reality". Of course, the same person may be examined again with regard to the same case. However, at this point the person is already changed by the original examination. (This naturally pertains only to a situation where the goals of the subsequent examination are identical to the goals of the original one; if the goals are different, the above comment does not apply.) In other words, the repetition of the examination is no guarantee of achieving the same outcome as in the original examination, since the very act of examination changes its subject. An expert in, let us say, chemistry, is in a different situation. If his/her task is to identify a certain substance, he/she may use a part of the available sample only, and another specialist may run tests on its remaining parts, regardless of the original testing.

{**DG 5**. This quality of PE was observed quite some time ago by polygraphists themselves. The issue must be noted, and its consequences kept to a minimum. However, it must not be blown out of proportion, as it is, for example, done by Vrij in the context of CQT tests. The title of his text alone (*A crucial* *and difficult role for the examiner: lack of standardisation in conducting the test*) is questionable. Moreover, the author's claim that "CQT examinations cannot be seen or presented as an objective and scientific process" (Vrij 2008: 309–311) is decidedly too far-reaching. Vrij disregards the existence of the standards ASTM E2062 and E2031 (PDD Examination Standards of Practice, Quality Control of Psychophysiologial Detection of Deception (Polygraph) Examinations). While it is true that through "speech pauses, tone of voice, voice loudness, etc." (p. 311) the outcome of a test may be distorted, the validity characteristics of CQT tests demonstrate that the situation is not quite that hopeless.}

A way to ensure at least partial inter-subjective controllability of PE is establishing a quality policy with regard to the performance of PE and subsequent quality control.

The issue of quality control with regard to forensic evidence is one of the most important issues contemporarily discussed in this science. Fundamental quality requirements of PE are as follows: the examination must be conducted before the subject is interrogated; the expert must undertake a careful analysis of the case and draft a set of test questions giving due regard to possible alternative courses of the event; a pre-test interview must be conducted in a non-accusatory manner, with the subject being offered freedom of expression; the examination must be conducted solely with the use of a test deemed to have been validated; the entire examination must be recorded and secured; the questions must be amended in cooperation with the subject; the polygrams must be evaluated with the use of validated numerical scoring systems. It is also necessary that all expert's notes or copies thereof be attached to the report. The report itself must include an account of the pre-test interview, a justification of the examination method selected, a presentation of calculations made, and the final opinion, which contains a clear statement (i.e. the final opinion may not use the term *probably* and similar terms, as illustrated by O1 and O2 above). Quality issues with regard to PE are regulated by the standard ASTM E 2031 (Quality Control of Psychophysiological Detection of Deception (polygraph) Examinations).

The level of inter-subjective controllability of PE is equal to the possible scope of control of examination quality, according to the quality policy in force in a given place and time.

The most important consequence of the above is that a single specific PE cannot be repeated under identical conditions, so the type of controllability that is characteristic for physics or chemistry is not applicable to it. Moreover, quality policy and its criteria are always arbitrarily established by someone who in this manner constructs a certain framework of the examinations which is constant neither through time nor, currently, through space.

Why are the issues of inter-subjective communicability and controllability so important? The answer is simple. Nothing else will place expert evidence as firmly within the realm of scientific statements. While inter-subjective communicability and controllability alone will not place evidence firmly within the realm of science, ensuring inter-subjective communicability and controllability will surely make it more worthy of the *scientific* descriptor.

IV

Knowledge used in expert evidence consists of generalisations drawn from scientific knowledge or otherwise based therein. Typically, a distinction is made between three types of generalisations: the "laws of science", well-grounded principles and the not-so-well-grounded (for the moment) research results (Anderson, Schum and Twining 2009: 270). In the empirical sciences, all these generalisations are idealistic in nature, or in any case must allow for certain exceptions. Consequently, no specific PE ever produces an absolutely certain result. This must be understood as follows: in a set of examinations performed, some outcomes (i.e. opinions) are true while others are false; yet it is impossible to tell, using only PE-related knowledge, which are which. In other words, the set of O-statements consists of true and false statements, and so does the set of F-statements.

The parameter that characterises the proportion of true and false statements is validity of the method used to determine the true/false value of F and O-statements. This parameter thus applies to the principles of issuing F-statements and principles of formulating opinions. Validity has two components. The first one is accuracy, which determines the distance between the outcome of the examination from the actual value of the variable under examination. The second one is precision, i.e. the degree to which the outcomes of examinations can be repeated. A method may be accurate but not precise, and *vice versa*, or it may be neither accurate nor precise. However, if a given method, according to certain criteria and in a certain context, is both accurate enough and precise enough, then this method is deemed justified, right, acceptable for the given needs, it is valid and its validity is known.

The process by which the accuracy and precision of a method is established, according to a set of criteria related to its forensic usefulness, is called the validation of the method. Sometimes, the process is referred to as developmental validation, as opposed to internal validation, which consists in establishing whether the outcomes generated with the use of a method in a specific laboratory conform to expectations (Tilstone 2000: 1309).

Reliability is a term related to the method, but also to evidence as a whole, and is a parameter that denotes the weight of evidence, which means that reliability needs to be considered in the context of a specific case. In the legal systems that list criteria for admissibility of evidence, reliability of evidence is discussed in the framework of certain criteria it should meet, for example the Daubert standard, on which more is forthcoming (Freckelton 2000: 715). Methodology distinguishes among various detailed aspects of reliability, e.g. the reliability of outcomes obtained with the use of a measuring instrument by one person, reliability of outcomes obtained with the use of a measuring instrument by differing persons, etc.

Determining the validity of various methods within forensic science is a highly specialised issue, and also a very complex one. Today, it is practically another branch of knowledge (Hadley, Fereday 2008; King, Maguire 2009; interestingly, difficulties in computing the rate of error are illustrated with PE-related examples also in Kaye, Bernstein and Mnookin 2004: §6.3.2).

The issue of the proportion of true and false statements in the F and O sets, or in fact the issue of validity of PE, has since the 1970s been the central problem under research surrounding PE. There is a vast amount of literature pertaining to this subject. At present, the matter is viewed somewhat differently, namely in the broader context of the so-called Daubert standard. The standard is formulated in the framework of law, but its main strength is in the discovery area, hence the enormous significance of the analysis of PE as scientific evidence.

The standard, sometimes referred to as the "scientific knowledge" approach, appears in various wordings. Let us cite here the following: "1. whether the theory or technique on which the testimony is based is capable of being tested; 2. whether the technique has a known rate of error in its application; 3. whether the theory or technique has been subjected to peer review and publication; 4. the level of acceptance in the relevant scientific community of the theory or technique; and 5. the extent to which there are standards to determine acceptable use of the technique" (Daniels 2002: 329).

The decision as to point 1 above belongs in fact to the realm of philosophy of science and constitutes the question whether the knowledge applied in PE belongs to science or, to use neo-positivistic language, to metaphysics. Three factors contributing to a positive answer (i.e. an answer in favour of the scientific status of PE knowledge) merit discussion. Firstly, since the sociological approach claims that science is what scientists do, it bears emphasising that issues of PE are taken up by noble respected institutions such as universities, and within these institutions by well-acknowledged academics of various specialisations. This is an argument that is hard to counter, since accepting that these academics put their professional efforts to something other than science would seriously subvert the current social order. Secondly, from a general methodological standpoint, the knowledge regarding PE has traceable origins, and the manner in which it was generated may be examined in view of methodology of empirical sciences, well accepted among psychologists, physiologists, etc. Thirdly, there is the specific methodological factor consists in verifying whether PE-related statements are inter-subjectively communicable and controllable; if so, they are testable by definition. That is precisely why Krapohl stresses so much the requirement of methodological accuracy of PE-related research when he writes: "[t]he research had to be published in full. (…) [t]he research had to be replicated" (Krapohl 2006: 150).

The issue from point 1 is directly related to the issue from point 3. Authors' publishing their examination results is a normal procedure, justified by reasons both academic and organisational, as well as good practice. Publications, naturally, differ in rank, and some of them undergo peer review. There is no reason to accept that only the suggestions of peer-reviewed publications should be incorporated into PE practice. A method is considered to be proper if its validity was established in the course of methodologically accurate research supported by other independent research, and this condition is deemed satisfied when the results of this research have been fully disclosed and published, and if the criteria for verifying the accuracy of decisions was independent of the outcomes of PE. Moreover, a test or technique should be possible to apply in typical real-life situations and assigned a name that allows for their unambiguous identification (Krapohl 2006). It is true that certain results obtained in highly advanced state-operated laboratories working in the national security sector are kept secret, at least for certain periods of time. On the one hand, there are the academic circles studying PE, and these circles are transparent. On the other hand, there is a realm of knowledge whose extent is difficult to establish and which fails to meet condition number 3. A realistic solution appears to be to assume that the knowledge from this realm is used in the course of covert operations conducted by the police and other services, and as such it never enters the justice system. Thus, studying this knowledge in the context of evidence is unnecessary.

Discussing point 4 is, essentially, taking a stand with regard to the Frye standard. There is no room in this paper to cite the entire literature on the issue – an issue burdened with difficulties and, fundamentally, impossible to decide clearly. It is impossible to tell who constitutes the relevant scientific community. Since ancient times the truth has been known that *nihil tam absurde* *dici posset, quo non dicatur ab aliquo philosophorum* (Cicero). On the other hand, there is no reason to arbitrarily disregard the views of any academic. Sometimes general acceptance in fact means a consensus among just a few individuals, not to mention other difficulties (Kaye, Bernstein, Mnookin 2004: § 5.3.3 b, c). At times, "general acceptance" was treated as a surrogate for validity (§ 5.3.2). There is research that shows that in various academic circles the degree of acceptance for PE as evidence rises as the level of knowledge about PE rises (Honts 2004: 114–115). Matte correctly distinguished between acceptance of PE as a general method and acceptance for specific parameters of the method and notes that "there is no doubt in the scientific community about the validity and reliability of the polygraph instrument currently being used in the field" (Matte 2000: 8). Actually, not accepting this claim would translate into rejecting a massive amount of scrupulously careful research on the validity and reliability of the method. Another author, on the other hand, accurately formulated the question with regard to general acceptance: "Is the theory upon which the hypothesis and/or technique is based generally accepted in the appropriate scientific community?", and offered a negative answer to this question with regard to CQT tests conducted both in the laboratory and in the field (Vrij 2008: 335). It is noteworthy how focusing on different aspects of general acceptance may radically alter the way in which the reader's opinion is shaped. Fortunately, the issue of determining the level of acceptance in the relevant scientific community is neither methodological nor epistemological, but rather sociological and political. The burden of finding a solution to it (*a casu ad casum*) is on the judge and his/her common sense, with the following in mind: "No standard for scientific evidence will *always* admit valid science and *always* exclude invalid science and pseudo-science. The choice between alternative forms of strict scrutiny must rest on the relative merits of the standards." (Kaye, Bernstein and Mnookin 2004: § 5.3.2).

The issue listed under (5), i.e. that of standardisation of expert methods, is one of the crucial issues in the assessment of expert evidence. I will attempt to prove that the existence of a standard applicable to an expert method constitutes one of the fundamental factors that allow for the inclusion of expert evidence in the realm of scientific evidence.

What remains is another issue of major significance with regard to PE, i.e. the issue of awareness of the rate of error in PE applications. As mentioned previously, the matter of diagnostic value of examinations used to hold great research interest. Nowadays, the overall effectiveness of specific techniques and tests within PE continues to be studies, but with the chief goal of deciding the admissibility of PE's use in specific situations. A method is admissible for evidentiary purposes if its validity is no less than 90%, and for investigative purposes if its validity is no less than 80%. These criteria are derived from the Daubert standard as well as the ASTM standards (Krapohl 2006).

Daubert, in its original form, was not exhaustive of all the legal demands towards expert evidence. There are discussions on post-*Daubert* validity factors, although they may at times be difficult to distinguish from methodological discussions. These factors are as follows: the expert should conduct research in his/her area of expertise independently of an expert opinion presented by him/her in a specific case; the expert opinion should maintain the same intellectual rigor as the research work; qualifications of the expert should be considered separately from the validity of eh method used, since the expert's qualifications, high as they may be, alone are never sufficient to demonstrate validity; the accuracy of the reasoning of the expert should be verified, and specifically, it should be verified whether the expert has not drawn an unjustified conclusion from a justified prerequisite (Kaye, Bernstein and Mnookin 2004: § 6.5.2).

The following consequences of the Daubert standard which are of significance for PE are also mentioned: clarity and coherence of the explanations offered by the expert pertaining to theories, methods and procedures; use of an expert who is independent from the parties to evaluate the original expert's opinion; and the expert's reputation (Dixon, Gill 2001: 38).

Let us then summarise briefly the answers to the *Daubert* questions with regard to PE. The answer to question 1 ("whether the theory or technique on which the testimony is based is capable of being tested") is affirmative. Contemporary knowledge regarding PE is being developed in accordance with the principles of methodology of empirical sciences and its high standards that are prevalent in psychology, physiology, medicine, sociology and other sciences. In this sense, the theories and techniques used in PE belong to the realm of hard science, although elements nearer to soft science may also be used.

{**DG 6.** Expert evidence that uses the instruments of natural sciences is sometimes referred to as hard-science based. In contrast, evidence based in psychology and social sciences is referred to as soft-science based. Initially this term was used to refer to psychological characterisations of individuals, chiefly perpetrators of crimes and their victims. It was stressed that sets of similar past events are the fundaments for issuing opinions; on their basis, forecasts were being made and assumed past behaviours reconstructed. This practice garnered a lot of reservations. The language used by psychologists was characterised as jargon or even as being esoteric. Moreover, reservations were expressed as to the practice where psychologists commented on the credibility of witnesses (Goodwin, Gurule 2002: 397 and next).

Probably any practising polygraphist would be able to cite examples of jargon or even esoteric language being used in PE. It certainly has an adverse effect on the image of PE among the general public. This only increases the importance of inter-subjective communicability in the area of PE.

Doubtlessly, PE-related knowledge includes and draws from general patterns of human behaviour, and as such it needs to have a "valid scientific foundation" (Kaye, Bernstein and Mnookin 2004: § 7.8.1).

It appears therefore that the division of the bases of evidence into hard and soft sciences belongs to the past. The latest literature pays little to no attention to it. *Khumbo Tire Co. Ltd. V. Carmichael* (1999) is of paramount importance here too, since it stretches the Daubert test to apply to any expert opinion (Henderson 2000: 725; National Research Council of the National Academies 2009: section 3–7)}.

The answer to question 2, probably the most important one ("whether the technique has a known rate of error in its application"), is also affirmative. Good practice recommendations allow only for validated methods and techniques to be used, ones with validity of no less than 80% for investigative applications and no less than 90% for evidentiary applications. Condition number 3 ("whether the theory or technique has been subjected to peer review and publication") is satisfied too: publications presenting research and examination results as well as theoretical analyses, all peer-reviewed, is available. Granted, not all PE-related publications meet this condition. Nonetheless, a person with an interest in PE, as long as they have a basic understanding of the structure of academic publications, may have access to publications compliant with condition number 3 and may verify against them the quality of a specific PE. The condition specified in number 4 ("the level of acceptance in the relevant scientific community of the theory or technique") is the most difficult one to satisfy. Aside from the comments offered above, one may rely on the practical usefulness of PE and whether the police and other services who use PE, as well as the teams of academics that support them, are in the process of carelessly wasting taxpayers' money. Condition number 5 ("the extent to which there are standards to determine acceptable use of the technique") must be considered relative to the time and place where the examination is conducted. In some countries, there are carefully constructed and practically applied PE-related standards, while in other countries no such standards are present. This applies to ASTM standards as well. In any case, standardisation of tests is possible, although the various stage of PE are standardisable to a various degree. Test formats are the easiest to standardise while the course of the pre-test interview is most difficult (on methodological difficulties in research of pre-test interview see: Horvath, and Buckley 2008). There is justified hope that the large-scale harmonisation processes that are taking place across forensic expertise will cover PE too.

How do these considerations on the Daubert standard impact the knowledge about PE? In the historical and developmental aspects, they do so very strongly. The Daubert case was a significant contribution to the development of science not only in the realm of PE but also in many other (or maybe even all) areas of forensic science. In the legal aspect, in the USA, i.e. the homeland of the Daubert standard, the impact seems to be considerably smaller: "Given the wealth of literature on the subject, an objective application of the *Daubert* factors to the polygraph should be quite capable of being accomplished both by experts in the field and by the judges who must make the ultimate determination of admissibility. The reality was, however, that *Daubert* did not result in opening the doors of American courts to use of polygraph evidence" (Daniels 2002: 330).

For us, the first of Daniels's sentences quoted is of importance. It is not a problem for PE experts to meet the conditions of the standard. Naturally, not all PE examinations do in fact meet these conditions; certainly practice demonstrates that many fail to do so, not to mention other quality problems. And yet, it is not these instances of shortcomings that form the model for aspirations.

If the conditions of the Daubert standard, in its broadest sense, are met, is this sufficient to consider PE scientific evidence? The answer must be negative. Passing the Daubert test (as some researchers call it) only opens up the possibility of discussion of the status of PE as scientific evidence. Let me generalise here. Daubert is a legal construct. It has a strong methodological undertone, but its character is normative, as is characteristic of the USA, where the law determines, *inter alia*, the conditions of admissibility of expert evidence. If we assume that this decision embodies the "scientific knowledge approach", then we can take it to be a model that could have theoretical application in other legal systems too. The reaction to Daubert outside of the USA suggests that this is indeed the case. Thus, we can accept that Daubert, together with its theoretical grounding, constitutes the best attempt at defining admissibility of scientific evidence currently available. The fundamental function of this standard is this very determination of admissibility. In fact, the role of this standard is gate-keeping. Only past this gate does the discussion begin on the actual characteristics of the evidence in question. The bare fact that the evidence was allowed past the gate in no way means that this evidence is scientific. I believe it is both critical and rational to claim that meeting the Daubert conditions

constitutes the methodological minimum for scientific evidence; it determines a threshold from which further debate may proceed. Therefore:
PE, if conducted correctly, satisfies the minimum methodological criteria to be considered scientific evidence.

V

If the question is: what is scientific evidence?, then there are two ways in which the answer to this question may be sought. The first one, explored above, is to define scientific evidence through admissibility criteria. This has proved to be insufficiently effective. The second one is to attempt an ostensive (deictic) definition, i.e. definition by illustration. In the present discussion, this would be accomplished by an indication of what is and what is not scientific evidence. The position of PE would be quite comfortable then. The fundamental literature on the subject counts PE as scientific evidence and analyses it as such (Goodwin, Gurule 2002; Imwilkenried 2004; Kaye, Bernstein and Mnookin 2004; Gianelli, Imwilkenried 2007). While superficially correct, the solution would be too modest for our purposes.
A reconstruction of an analytical definition, i.e. a definition that provides the meaning of a specific term in the language of a specific branch of knowledge (in our case, in the language of criminal procedure and forensic science), is not easy. Although the position of scientific evidence in today's criminal and civil procedure, the problem of "[w]hat counts as science, (…) who gets to make this decision, and how they should go about it are all hotly contested. Nor is this contest limited to the United States. The issue of scientific reliability is a hot topic in England and other Commonwealth jurisdictions, as well as in continental European systems" (Beecher-Monas 2007: 4).
In an attempt to distinguish between scientific evidence and other types of expert testimony, Kaye, Bernstein and Mnookin start out from the correct assumption that relying on philosophers of science is not the right approach. What is needed is more of a functional method which would enable such a distinction. The key here is the special or strict scrutiny that is necessary for the evaluation of certain forms of evidence: "Courts fear that it comes cloaked in an aura of infallibility that leads jurors to give it more credence than it deserves" (§ 5.2). However, this scrutiny must be applied to all and any expert evidence, and that is why, if "there is a rationale for a special rule for scientific expertise, it must be something special about science that justifies stricter scrutiny" (*ibidem*, §7.1). Possible issues include: difficulties in understanding the evidence, which is time-consuming and which generates costs; the persuasive force of

science, greater than with other forms of evidence, which may lead to overestimating and overvaluing the evidence; and the small number of experts capable of conducting an analysis of the scientific limitations of the evidence and of estimating the risk of error, which undermines trial guarantees such as cross-examination and opposing testimony. Only when these three prerequisites are met, is it justified to require heightened scrutiny and to elevate the evidence to the status of scientific evidence. To be precise, the authors write: “The court should consider whether these three concerns are present in sufficient degree to warrant heightened scrutiny” (*ibidem*). In courtroom practice, this may be an efficient criterion. It is however unproductive methodologically.

The importance of this condition, at the practical level, is nowadays being indirectly diminished. Dwyer claims that ultimately, all expert evidence may be understandable to a non-expert, since the basic tools of learning and understanding are equally accessible to specialists and non-specialists alike. If a non-expert is exposed to empirical information coming from the expert, then this information is based on the same principles that any other information in day-to-day life. Hence, the non-expert should be able to notice fundamental errors, if there are any, without the need to grasp the specialist concepts the expert may be using. At the fundamental level, commonsensical and scientific reasoning is the same. “The method of the specialist may be more refined, and so may in practice not initially be accessible to the non-expert (…). However, in principle a non-expert should have no difficulty in assessing the expert’s method, because he shares the same fundamental tools” (Dwyer 2008: 105). Strange as it may sound, refuting Dwyer’s claim would be tantamount to questioning the overall sensibility of using expert evidence within the justice system. Both views – the one just presented and the “heightened scrutiny” view – are able to be reconciled because the phenomenon of something being “not initially be accessible” constitutes grounds for heightened scrutiny.

In the courtroom practice, this may be a useful distinction. Methodologically however it is not fruitful. The point of the matter is that the expert’s opinion that is produced subsequently to PE requires interpretation. Let us consider once more the above-presented formulations:

O1a: Person (a_1) gave dishonest answers to relevant questions of the tests.
O1b: Person (a_1) gave honest answers to relevant questions of the tests.
O2a: The subject (a_2) recognizes a (certain) event.
O2b: The subject (a_2) does not recognize a (certain) event.
(Let us disregard for the time being O1c: Test of person (a_1) remains inconclusive.).

We assume that each of the examinations that led to O1a-O2b was conducted using one of the validated methods of PE, i.e. the validity of the method was at least 80%. What does this mean for a fact-finder from the evidentiary viewpoint in a criminal trial? In other words, how to evaluate this evidence? More generally speaking, which issues deserve heightened scrutiny?

Beecher-Monas (2007) proposes a framework for evaluation of scientific evidence. The work is very controversial and gives rise to numerous reservations. However, it merits discussion here, since it is quite up-to-date and is centred clearly on the same issues that are the focus of this paper.

According to Beecher-Monas, in order to conduct a correct evaluation of this type of evidence, it is necessary to perform the following five steps: "(1) identify and examine the proffered theory and hypothesis for their power to explain the data; (2) examine the data that supports (and undermines) the proffered theory; (3) employ supportable assumptions to fill the inevitable gaps between data and theory; (4) examine the methodology; and (5) engage in probabilistic assessment of the link between the data and the hypothesis" (Beecher-Monas 2007: 1).

Surprisingly, the author makes no mention of issues such as quality of the expert's opinion, accreditation of the laboratory, standardisation of examination methods, and qualifications of experts. Indeed, these matters are hardly included in the books at all. We will study them later in detail; for now, let us analyse the five steps proposed by Beecher-Monas.

The first two steps seem to be inseparable, since theory can hardly be examined without reference to the data that supports or undermines it. From the point of view of forensic science in general, and PE in particular, the concept of scientific theory as presented by Beecher-Monas is important. The author notes how the probabilistic reasoning is the foundation of contemporary science, and statistical laws its main product. (In the empirical sciences, *generalisations* is a better terms than *laws*, and we will use the latter here). Moreover, she observes that characteristics used in forensic identification "can be measured, and the frequency with which these characteristics appear in the general population is quantifiable" (*ibidem*: 41–47). To put it briefly, generalisations used as foundation of expert evidence are not exception-free. This is hardly a novel observation, but it bears repeating until forensic science is eventually free from the last remaining naïve notions of the exception-free, unique, individual or otherwise absolute status of the foundations of evidence.

However, there is a gap in Beecher-Monas's reasoning. The author focuses solely on the explanatory function of scientific laws and theories. She barely notes the descriptive function. And yet, all empirical learning begins with a description. Beecher-Monas ignores the issue of the optimisation task of scientific generalisations and the optimising statements themselves, i.e. the statements that answer the question: what should the situation be like or work like in order to achieve the desirable goals. In forensic science, the role of optimising statements is crucial. At times, it is less important to know the answer to the question, explanatory in nature, why something happens, and more important to know the answer to the question of how something should proceed, e.g. how to conduct a pre-test interview so as to achieve the aims assigned to this stage of PE. Here is an example of a PE-related generalisation:

(a) "(…) guilty subjects will produce larger physiological responses to the relevant questions to which they know they are deceptive, than to relatively unimportant comparison questions" (Honts 2004: 107).
For analytical purposes, statements of this kind are typically presented as implications. The edit then results in the following:
If the subject is a dishonest perpetrator, the subject's responses in PE area greater with regard to relevant questions than to comparison questions.
In a simplified edit with symbols:
$P(x) \rightarrow Q(x)$
For empirical reasons, a universal quantifier does not apply here, but an existential quantifier does.
This generalisation may be used for explanatory purposes, it may also be (and in fact is) used when formulating opinions in individual PE – naturally after being transformed into a more detailed form. We will soon consider how to cope with its probabilistic character.

(b) "…effective polygraph examinations necessitate the establishment of 'psychological set' from the outset of testing" (Holden 1997, following: Sosnowski, Wilcox 2009: 67–68).
This is not a descriptive statement. This is a technical directive that expresses an instruction as to what conditions must be met in order for PE to be effective. When attempting to bring a directive of this type to a more specific level, it is necessary to determine the desired effectiveness level and the factors that impact it and can be manipulated, at least to some degree. In the example cited, the point is to establish which factors impact the establishment of a 'psychological set' and how to manipulate them so as to achieve the goals of PE in an optimal manner.
As shown above, specific generalisations and optimising statements of importance for PE can be identified and analysed rather easily. The situation is hardly as comfortable when it comes to a general theory. Certainly there is a number of useful and interesting general-approach theoretical works (see e.g. Kleiner

2002; Kholodny 2006; Handler, Honts 2007). However, many theoretical aspects should be considered in the light of several alternative concepts, as is the norm with scientific endeavours.
The above ties in with step (2) of Beecher-Monas's list ("examine the data that supports (and undermines) the proffered theory"). The author observes that scientific theories arise out of many varying research attempts that at times lead to contradictory results, yet also lead to the synergy effect. It is the job of the expert to be able to reconcile the contradictions that arise (Beecher-Monas 2007: 47–48).

{**DG6.** Beecher-Monas uses this opportunity to criticise a court ruling: "The U.S. Supreme Court in *Joiner* made this mistake when, rather than assess how the various studies in conjunction supported the expert's hypothesis, it rejected each study seriatim as unable to justify the expert's causation conclusion." (*ibidem*: 47–48).
A large debate that took place in polygraphist circles illustrates this issue. It started with a text cited here above (Krapohl 2006). Matte offered a very critical response to it. Backster and Gordon expressed their positions in the debate as well. Krapohl published a response (see: *Polygraph* 2007, 36, 1). Matte continued presenting his views online at www.matte-polygraph.com. And yet, when a year later the same original position was presented at the seminar of the American Association of Police Polygraphists, appended with a broadened justification authored by Cushman, not one among the few hundred participants expressed any opposition. The position one takes with regard to the substance of the debate is irrelevant here. What matters is the illustration of the fact that an expert should be able to foresee situations of this type and to justify his/her view if needed, e.g. in court.}
The next step (3) suggested by Beecher-Monas ("employ supportable assumptions to fill the inevitable gaps between data and theory") results from the very nature of scientific knowledge, but not only that. Scientific knowledge does have gaps, and an expert may face the challenge of having to fill them in. Under such circumstances, one should use assumptions that, in the words of Beecher-Monas, "should be ensuring scientific validity; minimizing significant errors in estimating risk; maximizing incentives for safety research; and creating an orderly, predictable, and trustworthy process. Merely rejecting studies as "too speculative", as many courts dismissively call them, without explaining the basis for rejecting the underlying assumptions is not enough" (Beecher-Monas 2007: 51). We will consider the issue of "gaps between data and theory" in more details in a chapter to follow. For now, let me just make the following comments. What the author suggests is correct, although it appears that she

fails to make the clear distinction between a single examination and general research that I strove to outline in my initial epistemological declaration. The problem is that correct suggestions are not enough. The same suggestions are so general that, considering the varying level of expertise among experts, everyone may argue their position in court, whatever that position may be, using these very suggestions. What is needed is suggestions of much more detailed character. These however are nowhere to be found in the work of Beecher-Monas. And yet, such detailed requirements do exist, or least the ideological foundations for them do exist. I have in mind here the accreditation of laboratories, the standardisation of methods of conducting examinations, and the certification of experts. All of these will be dealt with here shortly.

One more issue related to the existence of gaps in expert knowledge must be taken into consideration. Namely, such gaps may serve as inspiration to undertake new original research.

{**DG 7.** I witnessed the following situation in the late 1970s. A corpse was found of a male who had fallen victim to very brutal murder, with indications of possible ritual murder. During the investigation, the police quickly apprehended three suspects. The suspects denied having participated in the crime and agreed to PE. During the pre-test interview each of them claimed to have no knowledge of any details of the event. However, in a series of POT tests run on each of the suspects, each subject revealed clear responses for the key questions. After the examination, all three suspects confessed. And yet during a hearing in court each of them withdrew the confessions, claiming that these had been forced out of them by the police. As to their knowledge of the details of the event, they testified that they had in fact had such knowledge (since "the whole town knew and it was in the papers"), but had been afraid to say so to the expert. The judge on the case asked the expert conducting the examination the following question: "if the subjects had in fact known these details from source other than having committed the act themselves, and if they had, for whatever reasons, claimed that they had no such knowledge in the course of the pre-test interview, would they have symptomatic responses for the key questions? The state of the art was such at the time that no clear answer to this question was known. Upon deliberation, the expert answered, honestly: "Theoretically, yes". The judge excluded the result of the examination from the evidence. The case was the impulse that resulted in an experiment (Konieczny, Frąś, Widacki 1984)}.

Step (4) ("examine the methodology") in Beecher-Monas's text makes no particularly valuable contribution to the discussions here. The author devotes as few as two pages to presenting the aspects of general methodology of empirical sciences that she finds important, and then segues from general principles of conduction scientific experiments directly into specific examinations. She then writes: “Courts in criminal cases have the most difficulty dealing with laboratory standard and protocols because criminal laboratories are largely unregulated and – with the notable exception of DNA evidence – **much of what passes for criminal evidence lacks any empirical support**” (p. 53, emphasis mine – JK). Yes, DNA evidence is the most methodologically advanced branch of forensic identification], but that hardly means that no problems occurred while it was being developed. And the comment that forensic (obviously not *criminal*!) laboratories “passes for criminal evidence lacks any empirical support”, offered without any argumentation to back it, cannot be treated seriously. It may have been intended to mean that the expert opinions generated by such laboratories have no connection to reality, and yet even Fred Zain, who generated countless fakes, did so to support the reasoning of the prosecution. Alternatively, it may have been intended to mean that no branch of forensic identification has good empirical and theoretical grounding. Surely the grounding they do have may be criticised, a changes in paradigm may be deemed necessary, and claims may be made that forensic sciences are undergoing a crisis, if the author so wishes. However, the stance that the author takes is not legitimate. Moreover, it is clearly ridiculous to claim, in 2007, that laboratories are “largely unregulated” in terms of standards and protocols. And so despite the issue’s overall utmost importance (“examine the methodology”), the way it is approached by Beecher-Monas fails to contribute to the refinement of PE. Step (5) (“engage in probabilistic assessment of the link between the data and the hypothesis”) in Beecher-Monas’s discussion seems to be similarly unpromising. Hence, this vital issue will be discussed here with reference to other works.

VI

The process of formulating opinions in PE involves a comparison between the intensity of the responses registered to the relevant questions with the intensity of responses registered to another type of questions, such as comparison or key questions, depending of the technique employed. It may be shown that this situation is typical of forensic evidence construed as an assessment of comparison. Consider the following: “The interpretation of scientific evidence may be thought of as the assessment of a comparison. This comparison is that between evidential material found at the scene of a crime (denote by M_C) and evidential material found on a suspect, a suspect clothing or around his environment (denote this by M_S). Denote the combination by $M = (M_{C,} M_{S.})$. (…) Qualities (…) or measurements (…) are taken from *M*. Comparisons are made of the source form and the receptor form. Denote these by E_C and E_S, respectively, and let $E = (E_C , E_S)$ denote the combined set. Comparison of E_C and E_S is to be made and the assessment of this comparison has to be quantified. The totality of the evidence is denoted by *E* and such that $E_v = (M, E)$ (Aitken, Taroni 2004: 16).

In the case of PE, material M_C is created in the mind of the person who participated in a certain event, for example in an act of committing a crime (possibly as the crime's perpetrator) and who wishes to hide the information related to this fact. Material M_S exists in the mind of any person. "Qualities" are constituted from aforementioned types of questions and responses to the questions, whereas "measurements" are the intensities of the responses. In the case of relevant questions we obtain E_C, and in case of other questions $-E_S$. Evidence from polygraph testing comprises: the questions used in the examination and the intensity of the responses registered after these questions (Konieczny 2007: 44).

Based on Champod and Evett (2009: 971–972) the following definition can be arrived at by means of careful interpretation: an expert opinion, which takes into account the entirety of circumstances related to the origin of the material under examination, which allows for a minimum of two hypotheses on the course of the event under examination, and which includes data on the probability of the outcomes, qualifies as scientific evidence. For an expert opinion like that, three principles apply: (i) scientific evidence must be interpreted within the context of the framework of circumstances, (ii) due consideration must be given both to the prosecution hypothesis and to the defence hypothesis (and, it appears, an initial assumption must be made that the two hypotheses are in opposition to one another; even if this assumption falls though, e.g. when the suspect confesses, both hypotheses must nonetheless be considered); (iii) the expert must answer two questions: what is the likelihood of the prosecution's position in the light of the given evidence, and what is the likelihood of the defence's position in the light of it. The ratio of these two likelihoods (i.e. the likelihood ratio, LR) is of crucial importance (*ibidem*).

Issues related to (i) have already been discussed here. Let us now move on to (ii) and (iii).

Let us consider the following to be the defence's hypothesis:

O1a: Person (a_1) gave dishonest answers to relevant questions of the tests.

The likelihood of this statement must be considered against the accuracy of the test on the basis of which this opinion was formulated. As mentioned previously, only the polygraph techniques that have accuracy above 90% may be applied for evidentiary purposes. Thus, only the Utah Zone Comparison Technique is an option here, which in deceptive cases gives 92% of accurate outcomes, after the exclusion of the opinion IC (Krapohl 2006: 149–152). We may then say that the likelihood of O1a is 92%. This translates into the following: In the set of N opinions of the O1a type, 92% of opinions are true; however, we do not know which ones have this quality of being true.

Let us now consider briefly the impact that various methods of interpretation of examination outcomes have on the final opinion. The leader among such methods is the numerical seven-point C. Backster scale, as modified: "Numerical scoring by adequately trained and experienced interpreters produces extremely high reliability that compares favorably with any psychological test interpreted by humans" (Raskin, Honts 2002: 21).

From among the computerised methods, there is most comparison between OSS (Objective Scoring System) and PA (probability analysis). In most general terms, OSS focuses on the differences in reactions to various types of stimuli and proceeds to integrate them mathematically. The system is being constantly developed; its third version OSS-3 is currently in use (Nelson, Handler, Krapohl).

PA uses the following decision-making protocol: if the probability of truthfulness of the subject (as indicated by the computer system) is no less than 0.70, then formulate opinion NDI, if it is less than 0.30, formulate opinion DI, and if it falls between these two values, consider the examination to be IC (Kircher, Raskin 2002: 307).

Both of these approaches have been declared valid and they constitute two acceptable methods of computing the data obtained in the course of PE (Webb et al. 2008: 254).

Other options also exist for computerised calculation of PE results. Depending on the methodology under which the comparison was carried out, the variance between results is smaller or larger. For instance, Gordon et al. 2006 compares two other algorithms alongside OSS, namely: ASIT Poly Suite and PolyScore 5.5 for one of the ZCT techniques. If IC outcomes were disregarded, the accuracy of the technique with all three algorithms was 100%.

Slavkovic proposes yet another manner of evaluation and analysis of polygraph data. She claims that her method is comparable, in terms of accuracy, with the results achieved with other computerised algorithms. Her work deserves attention also due to its meticulous treatment of methodological issues and its excellent intuition for the difficulties that are inherent in interpreting PE results (Slavkovic 2004).

Matte declared as early as 1996, after carrying out wide-ranging analyses, that there were no statistically significant differences between the results of calculations effected with the use of traditional scoring systems and those obtained using computer-based techniques (Matte 1996: 146).

Criswell, on the other hand, in 2007 "reported that the American Associate of Police Polygraphists has declared it unethical for an examiner to base an opinion solely on the results of a computer scoring algorithm" (after Nelson, Krapohl, Handler 2008: 210).

To recapitulate: the state of the art is such that it allows both the use of computer algorithms and the abstinence from their use in the evaluation of test results. The standard ASTM 2229 clearly permits this situation (Standard Practices for Interpretation of Psychophysiological Detection of Deception (Polygraph) Data). Best practice of course would entail using both of the methods – the numerical one and the computer one – and presenting the results of both in the report from the examination. This will only enforce the position of the expert opinion as scientific evidence. At the same time, one cannot forget that: "Just as polygraph testing cannot completely substitute for an adequate field investigation, computer algorithms cannot substitute for inadequately administered examinations that suffer from poorly selected examination targets, ineffective linguistic construction, or test data of inadequate interpretable quality. (…) As with any evaluation measure, ethical use of a test or automated process requires a reasonable understanding of its design, development goals, and operations, including its strengths and limitations" (Nelson, Krapohl, Handler 2008: 211).

{**DG8.** The theoretical background of the notion of LR and of the techniques of its calculation are presented in Champod, Evett 2009, as well as Aitken, Taroni 2004 and Kaye, Bernstein and Mnookin 2004: § 12.4.2. For the broad, forensic context see: Redmayne 2001, and in terms of general theory of evidence: Anderson, Schum and Twining 2009. This of course is just a very modest selection of research literature on the topic.}

The method of calculation of LR for PE was proposed by Kaye, who relied on an analogy to LR calculation in medical testing (Kaye 1987: 349; the paper was published in response to Lykken, Raskin and Kircher). The concept was later revisited and presented against a stronger background in Kaye, Koehler 2003: 349 and will be applied in the present paper to some extent.

Two parameters are of significance in medical testing: sensitivity and specificity.

Sensitivity, also known as the true positive rate, is the likelihood that the persons who suffer from a disease will be diagnosed correctly using the test. Specificity, also known as the false positive rate, is the likelihood that persons in the given population who are healthy will be diagnosed as suffering from the disease. Under these conditions:

LR = sensitivity/specificity

In terms of PE, we assume that for ZCT tests deception is the 'disease' and thus sensitivity is the likelihood of the correct indication of deceptive subjects, and specificity is the likelihood of an incorrect indication with regard to subjects who are non-deceptive, naturally within a certain population.

Let us use the examples sourced from Raskin, Honts 2002. The authors, having analysed the effectiveness of ZCT tests, write that in one study the true positive rate for guilty subjects was 92%, and the false positive rate, i.e. proportion of indications of deception among the innocent subjects, was 24%; inconclusive tests were disregarded. Reference is made here to Patrick, Iacono 1991, which includes <results from mean blind rescoring of the cases "verified with maximum certainty">. Thus:

LR = 0.92/0.24 = 3.9

Therefore, the hypothesis that the persons indicated as deceptive were indeed dishonest was in that case almost 4 times more likely to be true than the competing hypothesis.

In another study, cited in the same Raskin and Honts, sensitivity was 73% and specificity was 8%, the value of LR therefore was over 9. Both of these studies however were carried out on small samples and were not homogenous in terms of the structure of the study.

Nakayama (2002: 72) cites the results of a study using a much larger sample, with 1889 field cases. "The false negative rate was 2.5%", so the true positive rate was 97.5%, and the false positive rate was 0.4%. No specific information is provided on the exact techniques used; it is assumed that ZCT and/or CIT techniques were applied. The value of LR was then almost 244.

There is an abundance of examples to cite, and the list might be continued, followed by a demonstration of the disparities in LR values. The disparity is due to a variety of factors such as the size and characteristics of the population used in the study as well as the examination techniques used.

It is therefore possible to satisfy the conditions for the interpretation of scientific evidence set out in Champod and Evett, although it is not possible to do so consistently in every PE, at least not with full precision. If the expert is in possession of credible and methodologically correct data, including data on the population to which the specific PE pertained, then this expert is able to, and should, indicate the LR value in his/her opinion, or at least indicate the bracket within which this value falls. Ultimately, according to e.g. Kaye, "in judging whether polygraph results have substantial probative value" it is the LR value that is decisive (Kaye 1987: 361).

Apart from the LR method, another method too exists in PE which allows for the evaluation of evidence. This method, at least potentially, could have great impact on PE. The method consists in the construction and application of Bayesian networks (BNs), which formalises reasoning and enhances its precision under conditions of uncertainty, imprecision and similar difficulties. "A BN is a compact model representation for reasoning under uncertainty that formally combines elements of graph and probability theory. BNs allowing their user to define a pictorial representation of assumed probabilistic relationships among a set of variables, (…) are deemed to be relevant for a particular inferential problem." (Taroni 2009: 276). Benefits from using BNs include the following: the reasoning becomes more structured; it becomes necessary to take into account the evaluation of various options of the course of events; presentation of the inference is facilitated; the focus is on probability issues and their underlying assumptions (Taroni et al. 2004: 6).

It is beyond the scope of this paper to include a full example of the use of BNs in the context of PE. I will only cite here an example, inspired by the abovementioned paper, in which the application of BN would make excellent sense.

Let us assume that a crime has been committed and:

(i) there is a closed set of potential perpetrators, i.e. it is known with certainty that the name of the perpetrator is a part of a set where all the names are known, but it is not known which person on the list is the actual perpetrator; it is known that there is only one perpetrator,

(ii) PE was conducted on two persons from this set; both examinations had the DI outcome,

(iii) both validity and accuracy of the method used by the expert are known.

Under such conditions, how should one interpret the evidence obtained? The construction of a BN would enable an answer to this question and outline a clear route towards this answer. (Practically all parameters necessary to carry out the calculations are available here.) In practice, some cases are much less complex while other are much more complex, and the situation at times is not as comfortable, i.e. quantitative data is not available. This does not necessarily hinder the construction of a BN; good methodology is in place for the construction of such networks where data is missing (see: Biedermann, Taroni 2006).

Ultimately, we may then assume the following: it is possible to determine the approximate value of LR in the case of a specific PE, and it is also possible to apply BN in PE-related reasoning.

This statement sounds somewhat self-evident. Nonetheless, it is a message that bears saying clearly. It is also quite an evident conclusion that issues of LR and BN in PE merit further intensive research.

VII

The American National Research Council conducted a critical assessment of PE, the results of which were published in a report (The Polygraph and Lie Detection 2003). The first conclusion of the report is that "(...) in population of examinees (...), untrained in countermeasures, specific-incident polygraph tests can discriminate lying from truth telling as rates well above chance, though well below perfection" (p. 4). The claim is surprising in that it fails to specify what *perfection* it cites. F. Horvath argues: "In spite of the lack of scientific consensus about the accuracy of polygraph testing, it is seldom that the issue is placed in proper perspective. One important question to be asked is not how accurate polygraph is in the abstract, but rather how accurate it is relative to other types of evidence and other processes used to accomplish similar objectives. When considered in this light, the evidence is quite different from that usually presented. CQT polygraph testing in controlled conditions, for instance, shows an accuracy that equals or exceeds that of other common means of investigation. In one study it was shown that polygraph testing produced an accuracy that was comparable to results obtained by document examiners and fingerprint analysts, and greatly exceeded that of eyewitnesses. Equally important, polygraph testing was shown to have greater utility, that is, it was useful in more situations than other forms of similar evidence considered" (Horvath 2000: 1107); the issue of utility of PE will be raised here shortly. The study that Horvath mentions hereinabove is Widacki, Horvath (1978). Naturally, one might observe that Horvath's argument goes: 'sure, I am not perfect, but look, others don't look that good either'. Possibly this style of argument is better avoided. And yet, the facts speak for themselves, and 'the others' apart from not looking so good themselves, tend to be drastically arrogant too.

{**DG9.** J. Levinson, a specialist in questioned documents examination, quotes an opinion that he presented to the court, along with the resulting discussion – a model one, in his opinion, we are led to believe. The opinion ran along the lines of: In my professional opinion, document A was written by person B. The court asked if the document had been written by the accused. The answer was affirmative. The next question of the court was whether it was possible that a different person had written the questioned document. Levinson's answer was: "We can have philosophical discussions that <<anything is possible>>, but in real terms I have absolutely no doubt. The defendant wrote the Questioned writing, and no one else" (Levinson 2001: 187). Hats off to the author, who in those few words: (1) practically dismissed philosophy in its entirety, (2) put his own personal conviction before any other form of learning available to an expert, (3) smoothly put himself in the court's position or in the court's shoes, so to say, (4) introduced an interesting semiotic category ("real terms"), and (5) most likely, earned the prosecutor's gratitude.

Further in his publication, Levinson cites another opinion he delivered. The opinion was along the lines of: Document C was probably written by person K. The court asked what 'probably' meant in this context. In response, Levinson explained that he could not with absolute certainty exclude there being another author. The judge asked if that meant that there was a 60% chance of K being the author of the questioned material. The reply was: "I really do not want to express myself in numerical terms. My intention is, that I cannot totally exclude another writers who could have written the material. If I had significant doubts, I would have said <<K as well as other writers could have written the Questioned material>>" (*ibidem*). I doubt it merits a comment, even more so since the text was published a few years after the Daubert decision. Hopefully though this kind of comments forced and enhanced the realisation that forensic science was going through a deep crisis.}

The Polygraph and Lie Detection 2003 was strongly critical of the screening of candidates for the civil service and other positions in the public administration system: "Because actual screening applications involve considerably more ambiguity for the examinee and in determining truth than arises in specific-incident studies, polygraph accuracy for screening purposes is almost certainly lower than what can be achieved by specific-incident polygraph tests in the field" (p. 4).

These words may have rung true in 2003. If they did, then it must be noted that this criticism actually had an impact on the practice of PE. Methodology of screening examinations was updated to come into maximum alignment with the methodology of specific-incident testing. Recommendations suggest the use multi-issue tests in advanced CQT techniques, which practically eliminates the use of RIT and POT for screening purposes. The TES test was designed, and it comes with a set of quality requirements that are unrivalled in the history of PE. There are rigorously defined procedures for applications of "emergency" tests and clearing tests, which are interpreted with the used of the 7-point scale. It is also emphasised repeatedly that the outcome of PE may never constitute the sole ground for a decision in such a setting (Federal Psychophysiological 2004, Model Policy 2006, Model Policy 2010).

The counterintelligence services have been most interested in using PE in the interest of national security. Wettering reports on the possible uses of PE for such purposes and presents a noteworthy list of counterintelligence services' PE-related accomplishments. There are, of course, strong reasons to believe that the list is far being from exhaustive. Wettering notes PE's "accuracy rating of about 90%, which is comparable to fingerprint identification" (based on information from a justice system representative). He however goes on to say: "But American attitudes concerning civil liberties, distrust of government, and disrespect for secrets of any kind make mandatory routine polygraph examination only a pipedream of counterintelligence officers, not the reality." (Wettering 2009: 287). This sort of ideological background certainly makes it easier to be critical of PE.

Imwinkelried (2004) is another text abounding in important and interesting comments on critical approaches to scientific evidence. I would like to relate now some of the points Imwinkelried makes.

Imwinkelried (2004) lists the following areas which could be used to argue critically of an expert in court: "qualifications to vouch for the theory's validation, the theory's validity, the theory's general acceptance, (…) qualifications to attest the instrument's reliability, the instrument's reliability, and the general acceptance of the instrument." (p. 92). These comments are intended to apply to experts in general, if they are presenting scientific evidence. Some of them we have discussed here already with relation to PE. Let me recapitulate briefly: issues of validity of the polygraph as a scientific instrument are regulated by standards ASTM E2063 Calibration and Functionality Checks Used in Forensic Psychophysiological Detection of Deception (Polygraph) Examinations and ASTM 2439 Instrumentation, Sensors and Operating Software Used in Forensic Psychophysiological Detection of Deception (Polygraph) Examinations.

Imwinkelried specifies that what is crucial is not so much the general validity of the scientific branch the expert represents, but rather the validity of the specific theory or technique that was used as a basis for formulating the opinion (p. 96–97). This is a key issue for PE. The expert must be ready to provide answers related not only to the general questions regarding his/her specialisation, but also to questions that examine the particular techniques that he/she used, as well as the accuracy and theoretical background of these techniques. Moreover, the expert should be ready to present his/her academic accomplishments with regard to the issues of concern for the PE undertaken and to detail previous cases in which he/she previously appeared before courts in the capacity of expert (p. 98). Should the expert fail to demonstrate sufficient theoretical foundation, his/her opinion may be discredited by the court, on the grounds that the expert appears to be "a technician and law enforcement officer, not a scientist" (p. 99). This stretches the limits of understating the term *expert*, but seems instructive. A remedy to this type of problems might consist in appending the expert opinion with a bibliography of relevant research or even with copies of relevant scientific papers that present "solid empirical research" of relevance for the case (p. 108).

Generalities are therefore not sufficient. Imwilkenried continues: "For example, in the case of polygraph, the underlying theory is not that the polygraph can accurately detect and measure physiological changes. The question is not whether the polygraph can measure those changes with exquisite precision. Rather, the theory is that the analysis of those changes will enable the polygraphist to draw an accurate inference as to whether the subject is being truthful. That is the theory which must be empirically verified. As Justice Blackmun emphasized in *Daubert,* it is the expert's ultimate "inference" which "must be derived by the scientific method …[and] supported by appropriate validation." The dispositive question is the extent of the validation of *this* specific theory. Other, related theories may have been verified, but it would be a *non sequitur* to leap from their validation to the conclusion that the specific theory cited by the proponent's expert has been validated" (p. 108).

Imwilkenried points out that, for a person who opposes evidentiary use of PE outcomes, it is strategically beneficial to push the court towards a possibly broadest interpretation of "the relevant scientific circle", since it makes it easier to demonstrate that there opposition to PE exists and thus argue that the "general acceptance" condition is not met. "Polygraphy proponents argue that the courts should define polygraphy as the relevant scientific circle; since polygraphists believe in their own discipline, polygraphy should be admissible. (…) However, polygraphy's opponents have convinced courts that the relevant disciplines include the fields of psychiatry and physiology. (…) Since many scientists in those disciplines are sceptical of polygraphy, polygraphy is not "generally accepted" within the meaning of that expression in *Frye*." (p. 122).

Other authors, when commenting on PE from the legal standpoint, raise similar issues, with the possible difference in where emphasis is placed. Goodwin and Gurule (2002) discuss at length the issue of quality control standards, their existence and application (p. 284–286). Gianelli and Imwinkelried (2007) cite constitutional arguments for admissibility of PE (p. 466–470) and stress the issue of examination quality (p. 428–429).

While the legal issues are not at the heart of this paper, it is nonetheless not my intention to disregard the views of lawyers in general, and their input into the debate on methodological problems in particular. If the lawyers' comments were ranked in order of importance, the top of the list would go to these two claims: present a modern theory of PE, designed with input from psychologists, psychiatrists and statisticians; continue to standardise examination and develop relevant quality policy, together with methods of quality assurance and quality control. Issues of validity, reliability, accuracy, etc., important as they are in their own right, remain outside the very top of this list. This is because the progress that has already been made in these areas provides for a more optimistic outlook. It is noteworthy how fast methodological criticism of forensic polygraphy becomes outdated. This is evidence that polygraphists, as a group, are eager to learn and grow professionally, and to assume a self-critical approach. Polygraphy has come a long way, although, of course, a lot remains to be done.

{**DG10.** Gianelli and Imwinkelried (2007: 429) write about the case, famous a few decades ago, of Floyd Fay, who in 1978 was sentenced for murder – as it turned out, mistakenly. The defendant underwent PE, the outcomes of which were ten interpreted by five (!) experts who applied differing methods of evaluation of the polygrams. The first two experts declared the defendant deceptive. The third and fourth expert decided that the examination results were inconclusive. The fifth expert declared the defendant to be truthful. Moreover, a psychiatry professor testified in the case as well, and argued that the tests were of no value at all. Hopefully, today such a situation would not have occurred.}

VIII

Let us have another look at the statements that represent the opinions that conclude PE. Two will suffice for the present purposes:

O1a: Person (a_1) gave dishonest answers to relevant questions of the tests.
O2a: The subject (a_2) recognizes a (certain) event.

Both of these propositions effectively constitute identification, since they point out that an object belongs to a certain set. They also include an element of individualisation, since they assign this quality to specific objects, in this case to persons (a_1) and (a_2). The propositions invoke a certain classification, and "(…) all identifications are classifications. Some of the classes are simply larger than others. The larger the class, the less discriminating the identification, but all such associations provide relevant information" (Kaye 2009: 87).

Saks and Koehler (2005) opened a debate on the necessity of change in the paradigm of forensic sciences. Consequently, notions that had been in use traditionally, such as identification, individualization and uniqueness, came to be questioned. PE remained outside the realm of this discussion. The new outlook on the above-mentioned issues however may have an impact on PE, and it is worthwhile to consider what impact that may be.
Critical issues in this debate were the traditional areas in forensic examination, such as fingerprint examination, as well as examination of handwriting, tool marks, bite marks, hair (using microscope methods), bullets, shoe tracks and others, all of which share the central assumption that, if two marks/traces are indistinguishable, then they have been generated by the same object. Consequently, the second assumption arises: if two marks/traces have not been generated by the same object, they differ (note that the second assumption is by no means a logical conclusion of the first one). Additionally, there is also a third assumption, which consists in accepting that certain marks have discernible uniqueness, i.e. are unique, individual, special, and therefore allow the first two assumptions to come true (Saks, Koehler 2005: 892–895).
Saks and Koehler (2008) continue this critical approach to expert evidence based on the individualisation principle. The argument is as follows. Forensic identification consists in two steps. In the first step, one compares an item of evidence to an item gained from a known source and determines how similar they are. In the second step, one determines the likelihood of their common origin, as indicated above. Both steps are susceptible to errors, and in forensic science, the risk of an error indeed occurring is unknown. Practice reveals experts' errors; there were even cases where the same experts, examining the same material (unknowingly, after some time had passed) arrived at divergent conclusions, in all likelihood swayed by different information on the case context they were given (on purpose). Saks and Koehler argue that there is no scientific reason whatsoever to accept the individualisation principle. Even an expert who determines full similarity of evidence and the material offered for comparison should not give a categorical opinion on both having the same origin: "The expert should explain, that, in finding that two patterns match, they have placed the suspect object or person in a pool of one or more objects that match the evidentiary marks. The strength of the likelihood that the known object or person shares a common source with the questioned object or person depends on the size of the pool. No scientific justification exists for assuming that the size of the pool is one" (Saks, Koehler 2008: 216–217).
It has also been demonstrated that the traditional manner of formulating opinions, i.e. one that relies on the individualisation principle, works against the defendant and increases the propensity of a jury to find the defendant guilty. It

also limits the effectiveness of trial safeguard methods such as cross-examination, opposing expert and jury instruction (McQuiston-Surret, Saks 2008). Indeed, statements such as

Defendant (a_1) has written, in his/her hand, the questioned document (D)
or
The fingerprint (F_1) found at the crime was left by the defendant (a_1)

are highly persuasive and discourage argument. This is particularly true given the philosophical and methodological approach demonstrated by J. Levinson cited in the digression above. What is more, such opinions in fact take the court's place in making factual claims, and experts have no right whatsoever – moral or any other kind – to take that place. Experts' statements should read more along the lines of

Characteristics of the handwriting in document (D) are convergent with the characteristic of handwriting of the defendant (a_1)
or
The image of ridges in the fingerprint (F_1) found at the crimes scene is convergent with the image of ridges of one of the fingers of the defendant (a_1),

i.e. document (D) and fingerprint (F_1) are indeed "in the pool", but "[n]o scientific justification exists for assuming that the size of the pool is one".

For the sake of completeness of discussion, it should be noted that there are authors whose evaluation of the individualisation and uniqueness approach is not quite as critical. For instance, Kaye (2009) takes a moderate approach. Also, Redmayne, in the discussion of how fingerprint analysts formulate opinions, explains how the model used by the is that of "absolute identification that has been around for a long time and works relatively well." (2001: 51).

This brings us back to the original questions. The problem is as follows: are the statements such as "Person (a_1) gave dishonest answers to relevant questions of the tests" and "Person (a_2) recognizes a (certain) event" free from these problems?

{**DG 11.** These are hardly new issues for PE. The matter was given attention in United States *v.* Scheffer in 1996. When debating the admissibility of polygraph evidence, reservations were raised that there was "the danger that the opinion of the polygraph examiner will intrude on the jury's function of assessing credibility; (…) the danger that jurors will accord excessive weight to the expert's testimony; (…) the danger that the focus of the trial will shift from the guilt or innocence of the accused to the validity of the polygraph examination: (…)". The third reservation is not PE-specific; it may be made against any method of forensic examination. The first two reservations are important and have been thoroughly protested against. A full discussion of the case is presented in the special issue of *Polygraph* 1997, Vol. 26, No. 3, from which the above list is also cited (p. 137).

Overall, the debate is well summed up by the following statement: "studies refute the proposition that jurors are likely to give disproportionate weight to polygraph evidence", quoted in Daniels (2002: 330).}

The alternatives must be considered, and they are as follows. Instead of honesty, dishonesty and recognition of a certain event, it is possible to speak in terms of the occurrence (or lack thereof) of significant/symptomatic responses/reactions for relevant/key questions. A formulation of an opinion would then run along the lines of

In the course of PE, significant responses for relevant questions were observed in person (a_1).

where the actual questions asked would of course be included in the examination report.

This practice is acceptable and is even recommended in pre-employment polygraph screening examinations (Model Policy 2010). This is understandable in that screening examinations are usually "internal", performed within certain organisation and institutions, and the interpretation framework of outcomes such as "No Significant Reactions/Responses (NSR)" or "Significant Reactions/Responses (SR)" may be designed precisely with this organisation or institution in mind. However, in case of examinations conducted within a criminal or evidentiary investigation, this solution should never be applied, since an opinion thus generated would be incomplete. To be more precise, it would hardly be an opinion at all, but rather just a statement of facts, lacking in their interpretation and conclusions. Even if, in the course of such an examination, the need arises to conduct a single issue test, then "[u]nless prohibited by law, a professional opinion that an *examinee was deceptive*, based on physiological data, should only result from a specific issue test" (*ibidem*, p. 8). According to this document, professional opinions should therefore run along the lines of O1a and O2a.

It is however not so simple. Nothing is lost when employing statements such as

O1a': Person (a_1) reacted to the relevant questions in the test in the same manner as a person giving dishonest answers to these questions would react

or

O2a': Person (a_2) reacted to the key question in the same manner as a person recognizing a certain event would react.

Both O1a' and O2a' offer the same level of discovery. Yet in a sense they shift the responsibility for the consequences of the interpretation of the identification from the expert of the user, which is in line with the current tendencies in forensic science.

{**DG 12.** Statements such as O1a' and O2a' (or statements very similar thereto) were used in professional practice as early as thirty years ago by J. Widacki, a leading Polish specialist in forensic science and PE. The Polish courts were fully accepting of such opinions (Widacki 1982).}

Let us now discuss, briefly and very generally, the connection between forensic identification and the decision theory. The concepts of Biedermann, Bozza and Taroni (2008) are a good starting point for this discussion.

Let us assume that $\theta_1, \theta_2, \ldots, \theta_n$ symbolise certain states where likelihoods are, respectively, $Pr(\theta_1), Pr(\theta_2), \ldots, Pr(\theta_n)$, and the decision-making space, understood as action that may be taken in one of these states, is the (exhaustive and rozłączny) set $d_{1,} d_2, \ldots, d_m$. Making a specific decision, in a specific state, has consequences $C_{i,j}$. Each of the consequences has a specific utility denoted by the symbol u, i.e. $u(C_{i,j})$ (*ibidem*: 121).

A combination of states, decisions and their consequences may be illustrated as follows (I am paraphrasing here an example cited by the authors, translated into a PE-relevant context):

If the reality corresponds to the state θ_1 consisting in the subject giving dishonest answers to test questions, then the following situations are possible:

decision d_1 is made to choose identification (DI), with the consequence C_{11} (correct identification),

decision d_2 is made to declare the examination to be (IC), with the consequence C_{21} ("neutral"),

decision d_3 is made to choose identification (NDI), with the consequence C_{31} (incorrect exclusion).

If the reality corresponds to the state θ_2 consisting in the subject giving honest answers to test questions, then the following situations are possible:

decision d_1 is made to choose identification (DI), with the consequence C_{12} (incorrect identification),

decision d_2 is made to declare the examination to be (IC), with the consequence C_{22} ("neutral"),

decision d_3 is made to choose identification (NDI), with the consequence C_{32} (correct exclusion).

The above-listed consequences may be listed in order of utility:

$(C_{11}, C_{32}), (C_{21,} C_{22}), C_{31}, C_{12}$

The order of preference is descending from left to right, and the consequences in brackets have the same utility ranking.

The highest preference is awarded to the correct identification and correct exclusion, as they conform to the actual state, i.e. are true. The lowest ranking is awarded to false identification, and the second lowest to false exclusion, since this decision should never make the position of the subject after the examination worse than it was before; nonetheless, it is not false (*ibidem*: 123).

However, evaluation of utility in a specific context is not that straightforward. It depends on the likelihood of the state, and the likelihood is based on non-scientific circumstantial information and on the scientific evidence obtained. It also necessitates further assumptions and analyses (*ibidem*: 122 and 124). Ultimately, the final identification decisions are not made by the expert, or to put it more precisely: "(…) in a decision theoretic conceptualization of forensic identification, a recipient of expert information should not expect forensic scientists to be competent in providing informed and recipient tailored opinion regarding individualization (unless the scientist is given *and* uses his client's preferential system)" (p. 129).

Very few actual calculations based on this method exist, which the authors admit themselves. Medical sciences are an exception. I will attempt below to provide an example, paraphrased from an example offered by the authors Taroni, Bozza and Aitken (2005: 897).

Let us assume that the decision-maker in a trial is supposed to take action, selecting from specific courses thereof, in specific states that are unknown to him/her. Each option is assigned a utility rating on a scale from 0 to 10 (please note that the actual numbers are assigned almost at random). The decision-maker in the trial has got certain grounds that lead him/her to consider arresting a suspect. He/she however continues seeking further grounds before proceeding, e.g. scientific evidence. The analysis develops as follows:

(i) arrest the suspect if the suspect is innocent – utility ranking 0 (zero),
(ii) arrest the suspect if the suspect is guilty – utility ranking 10,
(iii) not arrest the suspect if the suspect is innocent – utility ranking 4,
(iiii) not arrest the suspect if the suspect is guilty – a utility ranking 0 (zero).

Let us further assume that the decision-maker decides to conduct PE, and the suspect conceded to it. After the examination is completed, the expert issues an opinion disadvantageous for the suspect (DI).

Let us proceed to design a diagram where on the x (horizontal) axis we mark the accuracy values of the method of identification employed, in the range from zero to 1, and on the vertical axes cutting through points 1 and 0 of the accuracy marking we mark the utility values and join them with straight lines, (i) with (ii) and (iii) with (iiii). The junction of the charts enables the reconstruction of the minimal value of accuracy, starting from which the decision to arrest the suspect becomes optimal. In the example outlined above, this value is approximately 0.7. This translates into the following: if the test used by the expert in PE had validity/accuracy greater than that number, then the decision to arrest the suspect is indeed optimal, given the utility values as assigned in the example. Further details of application of this method may be found in the two above-referenced texts.

Clearly then the final decision on individualisation should not and must not be taken by the expert. The decision belongs to the recipient of expert information and it is this recipient that is responsible for it.

IX

Let us now attempt to arrive at a specific concept of scientific evidence. To say that expert evidence is included in the category of scientific evidence on the basis of the method used by the expert, and even to add that this method should be scientific, is not saying much. Firstly, this encourages the misleading impression that the discovery-oriented activities performed by the expert constitute science, which we have already said is not the case. Secondly, it invokes a problem similar to (or even identical with) the problem of "general acceptance", with all its attendant difficulties.

Let us then agree that within forensic science, a scientific method is a manner of conduct that is in accordance with the current standard of examination in a given branch of science. This approach is susceptible to criticism on the grounds that it is that very branch of science that determines the current standards of expert conducts. It is so indeed, and yet: (i) an expert is not a scientist, or at least not every expert is a scientist; (ii) the expert's duty of continuing professional education is not equivalent to the duty to monitor state-of-the-art developments of his/her relevant branch of science to the degree that would allow immediately updated implementation decisions to be made in practice, since (iii) making such decisions requires more than generic expert knowledge, and in particular it requires an understanding of the broad context of forensic science, as well as organisational, economic and legal circumstances. This is why, with few exceptions, discovering a new scientifically sound method, or outlining it in an adequately rigorous manner, is separate from the decision to implement this method in practice. Kiely puts it succinctly in the question: "How does the criminal discovery provide for the exchange of scientific information between the prosecution and defense?", followed by: "The first big subject involves the question of what the appropriate standards of "forensic" science are that can support a proffer fact that can be used to establish a material fact in case" (Kiely 2001: 44).

Let us agree then that scientific evidence is a proposition of an expert that is inter-subjectively controllable, that is grounded in an examination, and that is conducted in accordance with the current standards and principles of best practice in a given branch of science.

Let us now attempt a characterisation of PE, taking that definition as a starting point.

Generally, evidence is defined as information that may be presented in court, in order to determine the likelihood of certain facts relevant for the case, in such a manner that this information may prove or disprove these facts (Johnstone, Hutton 2009: 93). In other words, evidence is an item of information that justifies further information (Stein 2005: 35).

Information as such is therefore not yet evidence; it becomes evidence when it gains the capacity to be used as a premise in reasoning relevant under the law. Twining (2006: 438) notes that this understanding of information is shared by all theoretical approaches to evidence.

For instance, expert A presents to the authority involved in trial proceedings an item of information, such as for example O1a'. This information is not evidence yet. It will qualify as evidence only after it has been used in a reasoning of the following type: Expert A claims that O1a', **and therefore** O1a'.

If we agree that information is the correct *genus* for evidence, then we must consider what exactly constitutes its medium. Statement O1a' is a sentence from the syntactic point of view, and the content of this sentence, i.e. the proposition, is the medium of information. This is only true on the condition that the sentence is understandable, i.e. carries meaning. This in turn is one of the conditions for inter-subjective controllability. (Naturally, whether the proposition is understandable is a matter fully separate from whether this proposition is true or false.) The sentence must carry *meaning*, singular, not *meanings*, plural; if an opinion is ambiguous in meaning, it will not be inter-subjectively sensible.

We have thus far demonstrated that (i) the outcome of PE is information, the medium of which is a proposition understood as the content of an opinion, and that (ii) this information may be used as a premise in reasoning relevant under the law.

Whether the full condition of inter-subjective controllability has been satisfied may be determined on the basis of a full PE report. This condition is also a *sui* *generis* bridge between the science and the practice of formulating opinions, since it is an aspect of the quality of being scientific that is clearly accessible in the discovery process of PE.

{**DG 13.** This is an area somewhat susceptible to problems connected with the expert's reference to his/her professional experience. This experience may involve knowledge that is intangible, subtle, difficult to evaluate and at times impossible to question. Should the expert testify, for example: "I looked up at the stars and on this basis I claim that O1a", the situation would be clear. If on the other hand the expert testifies: "My professional experience indicates that…", or "Practice demonstrates that….", the matter of inter-subjective controllability becomes more convoluted. I will address this problem again shortly.}

Standardisation in the area of scientific evidence requires a discussion of the following assumptions: (1) there is a creator of standards, (2) the creator of standards is rational, i.e. prefers the solutions from within the pool of state-of-the-art knowledge that are most beneficial for the creator, (3) the creator of standards believes that output quality is fundamental, (4) the creator of standards is familiar with state-of-the-art knowledge, (5) an efficient mechanism is in place to motivate the expert to observe the recommendations of the creator of standards.

Condition (1) is obvious but complex, as it requires a decision as to where, who, and how, under given political, legal and cultural circumstances, established the creator of standards; the creator of standards may be self-established. Condition (2) assumes that the creator of standards understands that promoting trashy intellectual quality undermines its very existence. Condition (3) comes with the same consequence. It is here that the room for manoeuvre of the creator of standards is the greatest; the matter is contentious whether and to what extent standardisation should apply to accreditation of laboratories where research, training and certification are conducted. Condition (4) guarantees that state-of-the-art developments make their way into the standards, as well as promotes updates of recommendations. Condition (5) is the condition on which the entire point of standardisation depends.

Declarative endorsement of conditions (1)-(5) does not automatically equal success. The creator of standards may only believe itself to be familiar with state-of-the-art knowledge, and yet its political connections may make it possible for such a creator of standards to continue operating. Funding shortages may be an obstacle to implementation of good standards, as may be organisational inertia and plain unwillingness to develop professionally on the part of expert institutions. Lobbying on the conservative side (in the negative sense of the term) is ample in such situations. Finally, there may be experts ("experts"?) who will raise the flag of independence, in what is always an effective move, and boycott the standards, usually for ulterior motives.

Best practice recommendations may be approached as complementary to standardisation. Standards must be stable in order to be useful, and must therefore only be updated at reasonable intervals. However, scientific and practical progress is often incremental, innovative rather than revolutionary. These are often small but important amendments to practical solutions. Let us consider an example. ASTM Standard E 2229-02 (Standard Practices for Interpretation of Psychophysiological Detection of Deception (polygraph) Data) allows for admissibility of polygram interpretation using the global evaluation method, and outlines the general conditions for the application of this method. It also adds: "When possible, numerical evaluation shall be preferred over global evaluation" (section 4.1.2). Recent years have brought many modifications and alterations in numerical evaluation. Nonetheless, some experts insist – and have insisted for decades now – on using global evaluation methods only. While it would be difficult to argue misconduct in their case, they are definitely not implementing best practice. New solutions are constantly being developed; recent ones include intertest stimulation and two-stage examination, with special updates introduced in numerical scoring. Best practice recommendations are drafted by reputable experts: institutions, professional organisations and outstanding individuals. *Polygraph* has been publishing a series under the general title *Best Practices*. Interestingly, in the newest text in the series (Krapohl 2010), the author makes the reservation: "The opinions in this article are those of the author, and do not necessarily represent those of the US Department of Defense or Government" (p. 124).

The expert's professional experiences must be considered in two aspects. Firstly, experience is valuable, and an experienced expert may enjoy a reputation for being reliable, credible, and sometimes even cherished by the justice system. In very particular cases, professional experience may burden the expert to some extent. However, such problems tend to be minor and easily overcome.

{**DG 14.** For example, at the outset of my own experience with PE, I learned from more advanced colleagues that a subject who, in the course of the pretest interview, requests a glass of water or permission to smoke a cigarette, is then always evaluated as (DI). Further experience confirmed this "generalisation". Such notions may of course never be used as grounds for formulating opinions. It is however impossible to say with absolute certainty that they remain without effect on the expert's subconscious mind.}

A different and very dangerous situation occurs when an expert formulates an opinion solely on the basis of his/her personal experience, self-training and individual understanding of his/her area of specialisation. Haber and Haber (2008) analyse such a situation with regard to fingerprint examination, and come to the following conclusion: "So long as fingerprint examiners claim a scientific status for their ACE-V work product, assert perfect or near-perfect accuracy and fail to provide evidence of validity, standardized training, adequate proficiency testing and quality controls, we and other scientists will continue to ask the courts to exclude fingerprint comparison conclusions until evidence of their accuracy is scientifically demonstrated" (p. 149). If we swap *polygraph examination* for *fingerprint examination*, we must concur with the above.

A certain synthesis of the theoretical foundations of PE, standardisation requirements and best practice is accomplished in the case assessment and interpretation model (CAI), a new approach focused on transforming the achievements of forensic science and its applications for criminal justice. CAI originated from the necessity to provide answers to important questions concerning output quality, as well as an increase in requirements, caused by the scarcity of resources available for obtaining forensic evidence as well as the need to set new directions for the development of expert services (Jackson, Jones 2009).

All practising experts know how difficult it is at times to establish what the police or prosecutors in charge of the case actually want, what is the purpose of the PE they order, what the content of the test questions should be, what the expectations are with regard to the upcoming PE, etc. CAI provides the answers, by means of outlining the following stages of the model: "define the customer requirement; assess how forensic science can help; agree on a case examination strategy; carry out examination; interpret the results; and communicate the test results and opinions" (*ibidem*: 489).

These suggestions may seem trivial, but they are far from that. Moreover, Jackson and Jones offer, appended to their text, a thorough checklist for CAI which includes questions such as: "What strength of evidence is required – for charging or for prosecution ? (…) Are the issues investigative and/or evaluative? What level in the hierarchy of issues (subsource/source/activity/offense) will you be addressing? What type of opinion (explanation/posterior probability/single likelihood/likelihood ratio) will you be offering? (…) Ensure the strengths and limitations of your opinion are clearly set out and understood by the customer" (p. 495–496).

There is no doubt that CAI must be adopted for PE purposes urgently.

Let us return once more to the statement: Expert A claims that O1a', **and therefore** O1a'.

If the decision-maker in the trial accepts this inference, he/she thus includes proposition O1a' in the evidence in the case. However, the statement makes no claim as to whether the relevant fact actually occurred; it is a probabilistic statement. This is the key "(...) and difficult inferential problem that arises when the evidence in probabilistic reasoning is singular, unique, or one-of-a-kind" (Schum 2000: 587). In order to affect that inference, D.A. Schum introduces the notion of "ancillary evidence" or "meta-evidence", which is intended to assist in evaluating the believability of the inference. Meta-evidence consists chiefly in examining the generalisations that form the foundation of reasoning (p. 592–593; see also: Schum 2009). In our case, it might be a statement along the lines of: Whenever in the course of PE such and such physiological reactions have been observed, the subject was *probably* providing dishonest answers to the relevant questions of the tests. From the theoretical standpoint, a detailed and multifaceted analysis of this statement is possible. There is however no reason why meta-evidence should be limited to theoretical foundations of formulating opinions only. One may and one should include in this category the broadly conceived issues of examination quality, proper CAI implementation and – with equal emphasis – standardisation of methods used, accreditation of research institutions and certification of experts. Twining (2003) is most certainly right in the claim that "evidence is a multidisciplinary subject".

X

We are thus transported to the heart of the matter. It is the analysis of meta-evidence that is crucial in determining whether evidence is or is not scientific. This paper presents, to a modest extent, the criteria that may be applied in such analysis with regard to PE.

Where then is the borderline between scientific and non-scientific evidence? The answer is: no such borderline exists. Instead, there is a broad frontier along which more and less scientific evidence is situated. The quality of being scientific is gradable, incremental. Just as we may say that one theory is more scientific than another, so too we may say that one PE is more scientific than another. The point is for the more scientific ones to abound.

There is however no sure-fire guarantee of success. On occasion (hopefully a rare one!), an "expert" will be right whose examination will only consist in "looking up at the stars and claiming that O1a'", and a specialist will be wrong, despite best scientific efforts. Such, however, is the nature of science, because such is the nature of the universe.

Bibliography

C. Aitken, F. Taroni (2004), *Statistics and the Evaluation of Evidence for Forensic Scientists*, John Wiley & Sons, Chichester,

T. Anderson, D. Schum, W. Twining (2009), *Analysis of Evidence*, Cambridge University Press, Cambridge,

E. Beecher-Monas (2007), *Evaluating Scientific Evidence. An Interdisciplinary Framework for Intellectual Due Process*, Cambridge University Press, New York,

A. Biedermann, F. Taroni (2006), *Bayesian Networks and Probabilistic Reasoning about Scientific Evidence when there is a Lack of Data*, Forensic Science International, 157, 163–167,

A. Biedermann, S. Bozza, F. Taroni (2008), *Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications*, *Forensic Science International,* 177, 120–132,

C. Champod, I.W. Evett (2009), *Evidence Interpretation: A Logical Approach*, (in:) A. Jamieson, A. Moenssens (eds.), *Wiley Encyclopedia of Forensic Science*, Wiley, Chichester,

A. L-T Choo (2006), *Evidence*, Oxford University Press, Oxford,

C.W. Daniels (2002), *Legal Aspects of Polygraph Admissibility in the United States*, (in:) M. Kleiner (ed.), *Handbook of Polygraph Testing*, Academic Press, San Diego,

L. Dixon, B. Gill (2001), *Changes In the Standards for Admitting Expert Evidence In Federal Civil Cases Since the* Daubert *Decision*, RAND Institute for Civil Justice, Santa Monica,

D. Dwyer (2008), *The Judicial Assessment of Expert Evidence*, Cambridge University Press, Cambridge,

Federal Psychophysiological Detection of Deception Examiner Handbook (2004), Department of Defence, Washington,

I. Freckelton (2000), *Evidence*, (in:) J.A. Siegel, P.J. Saukko, G.C. Knupfer (eds.), *Encyclopedia of Forensic Sciences*, Academic Press, San Diego,

P.C. Gianelli, E.J. Imwilkenried (2007), *Scientific Evidence*, LexisNexis, Newark–San Francisco,

R.J. Goodwin, J. Gurule (2002), *Criminal and Scientific Evidence. Cases, Materials, Problems*, LexisNexis, Newark–San Francisco,

N. J. Gordon, F. B. Mohamed, S.H. Faro, S.M. Platek, H. Ahmad, J.M. Williams (2006), *Integrated Zone Comparison Polygraph Technique Accuracy with Scoring Algorithms*, Physiology & Behavior, 87, 2,

L. Haber, R.N. Haber (2008), *Experimental or Scientific Expertise*, Law, Probability and Risk, 7, 143–150,

K. Hadley, M.J. Fereday (2008), *Ensuring Competent Performance in Forensic Practice. Recovery, Analysis, Interpretation, and Reporting*, CRC Press, Boca Raton,

M.D. Handler, C.R. Honts (2007), *Psychophysiological Mechanism in Deception Detection: A Theoretical Overview*, Polygraph, 36, 4,

C. Henderson (2000), *Expert Witness*, (in:) J.A. Siegel, P.J. Saukko, G.C. Knupfer (eds.), *Encyclopedia of Forensic Sciences*, Academic Press, San Diego,

H.L. Ho (2008), *A Philosophy of Evidence Law. Justice in the Search of Truth*, Oxford University Press, Oxford,

E.J. Holden (1997), *Understanding Pre- and Post-conviction Polygraph – The Key to its Successful Use: A Guide*, Voice for the Defense, 26,

C.R. Honts (2004), *The Psychophysiological Detection of Deception: An Applied Psychology with a Long History and Wide Application*, (in:) P.A. Granhag, L.A. Strömwall (eds.), *The Detection of Deception in Forensic Contexts*, Cambridge University Press, Cambridge,

F. Horvath (2000), *Lie Detection*, (in:) J.A. Siegel, P.J. Saukko, G.C. Knupfer (eds.), *Encyclopedia of Forensic Sciences*, Academic Press, San Diego,

F. Horvath, J. Blair, J. P. Buckley (2008), *The Behavioral Analysis Interview: Clarifying the Practice, Theory and Understanding of its Use and Effectiveness*, European Polygraph, 2, 2(4),

E.J. Imwilkenried (2004), *The Methods of Attacking Scientific Evidence*, Lexis-Nexis, Newark–San Francisco,

G. Jackson, P.J. Jones (2009), *Case Assessment and Interpretation*, (in:) A. Jamieson, A. Moenssens (eds.), *Wiley Encyclopedia of Forensic Science*, Wiley, Chichester,

K. Jaegermann (1991), *Opiniowanie sądowo-lekarskie. Eseje o teorii*, Wydawnictwo Prawnicze, Warszawa,

D. Johnstone, G. Hutton (2009), *Blackstone's Police Manual*, vol. 2: *Evidence & Procedure*, Oxford University Press, Oxford,

C.A.G. Jones (1994), *Expert Witnesses. Science, Medicine, and the Practice of Law*, Clarendon Press, Oxford,

D.H. Kaye (1987), *The Validity of Tests: Caveant Omnes*, Jurimetrics Journal, 27,

D.H. Kaye, J.J. Koehler (2003), *Misquantification of Probative Value*, Law and Human Behavior, 27, 6,

D.H. Kaye, D.E. Bernstein, J.L. Mnookin (2004), *The New Wigmore: Expert Evidence*, Wolters Kluwer Law & Business, Austin,

D.H. Kaye (2009), *Identification, Individualization and Uniqueness: What's the Difference?*, Law, Probability and Risk, 8, 85–94,

Y. Kholodny (2006), *Interrogation Using Polygraph and its Scientific Basis*, Polygraph, 35, 1,

T.F. Kiely (2001), *Forensic Evidence: Science and the Criminal Law*, CRC Press, Boca Raton,

W. King, E. Maguire (2009), *Assessing the Performance of Systems Designed to Process Criminal Forensic Evidence*, Forensic Science Policy & Management: An International Journal, online downloaded August 2010, S.F. Kircher, D.C. Raskin (2002), *Computer Methods for the Psychophysiological Detection of Deception*, (in:) M. Kleiner (ed.), *Handbook of Polygraph Testing*, Academic Press, San Diego,

M. Kleiner (2002), *Physiological Detection of Deception in Psychological Perspectives: A Theoretical Proposal*, (in:) M. Kleiner (ed.), *Handbook of Polygraph Testing*, Academic Press, San Diego,

J. Konieczny, M. Frąś, J. Widacki (1984), *Specyfika tzw. śladu emocjonalnego a wybrane cechy osobowości w badaniu poligraficznym*, Archiwum Medycyny Sądowej i Kryminologii, 34, 1,

J. Konieczny (2007), *Selected Aspects of Uncertainty in Polygraph Examination*, European Polygraph, 1, 1,

D. Krapohl, S. Sturm (1997), *Terminology Reference for the Science of Psychophysiological Detection of Deception*, American Polygraph Association, Chattanooga,

D. Krapohl, J.B. McCloughan, S.M. Senter (2006), H*ow to Use the Concealed Information Test*, Polygraph, 35, 3,

D. Krapohl (2006), *Validated Polygraph Techniques*, Polygraph, 35, 3,

D.J. Krapohl (2010), *Short Report: A Test of the ESS with Two-Question Field Cases*, Polygraph, 39, 2,

J. Levinson (2001), *Questioned Documents: A Lawyer's Handbook*, Academic Press, San Diego,

D.T. Lykken (1981), *A Tremor In the Blood. Uses and Abuses of the Lie Detection*, McGraw–Hill Book Company, New York,

J.V. Matson, S.F. Daou, J.G. Soper (2003), *Effective Expert Witnessing*, CRC Press, Boca Raton,

J.A. Matte (1996), *Forensic Psychophysiology Using the Polygraph. Scientific Truth Verification – Lie Detection*, J.A. M. Publications, Williamsville–New York, J.A. Matte (2000), *Examination and Cross-Examination of Experts in Forensic Psychophysiology Using The Polygraph*, J.A. M. Publications, Williamsville–New York,

D. McQuiston-Surret, M. Saks (2008), *The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear*, Law and Human Behavior, on line, March 2009,

Model Policy (2006), *American Polygraph Association Model Policy for Law Enforcement Pre-Employment Polygraph Screening*, Polygraph, 35, 3,

Model Policy For Law Enforcement/Public-Service Pre-employment Polygraph Screening Examination, American Polygraph Association, www.polygraph.org, August 2010,

A. Moenssens (2009), *Expert Witness: Who Is?*, (in:) A. Jamieson, A. Moenssens (eds.), *Wiley Encyclopedia of Forensic Science*, Wiley, Chichester,

M. Nakayama (2002), *Practical Use of the Concealed Information Test for Criminal Investigation in Japan*, (in:) M. Kleiner (ed.), *Handbook of Polygraph Testing*, Academic Press, San Diego,

National Research Council of the National Academies (2009), *Strengthening Forensic Science in the United States: A Path Forward*, The National Academies Press, Washington,

R. Nelson, M. Handler, D. Krapohl, *Development and Validation of the Objective Scoring System*, version 3, *Lafayette Instrument's OSS-3 Report*, www.oss3.info, September 2010,

R. Nelson, D.J. Krapohl, M. Handler (2008), *Brute-Force Comparison: A Monte Carlo Study of the Objective Scoring System version 3 (OSS-3) and Human Polygraph Scorers*, Polygraph, 37, 3,

C.J. Patrick, W.G. Iacono (1991), *Validity of the Control Question Polygraph test: The Problem of Sampling Bias*, Journal of Applied Psychology, 76,

[The] Polygraph and Lie Detection (2003), The National Academies Press, Washington,

D.C. Raskin, C.R. Honts (2002), T*he Comparison Question Test*, (in:) M. Kleiner (ed.), *Handbook of Polygraph Testing*, Academic Press, San Diego,

P. Roberts, A. Zuckerman (2004), *Criminal Evidence*, Oxford University Press, Oxford,

M.J. Saks, J.J. Koehler (2005), *The Coming Paradigm Shift in Forensic Identification*, Science, 309, 5736,

M.J. Saks, J.J. Koehler (2008), *The Individualization Fallacy in Forensic Science Evidence*, Vanderbilt Law Review, 61, 1,

D.A. Schum (2000), *Singular Evidence and Probabilistic Reasoning in Judicial Proof*, (in:) J. F. Nijboer, W.J.J.M. Sprangers (eds.), *Harmonization in Forensic Expertise. An Inquiry into the Desirability of and Opportunities for International Standards*, Thela Thesis, Amsterdam,

D.A. Schum (2009), *A Science of Evidence: Contributions from Law and Probability*, Law, Probability and Risk, 8, 197–231,

A.B. Slavkovic (2004), *Automated Scoring of Polygraph Data*, (in:) H. Bozdogan (ed.), *Statistical Data Mining and Knowledge Discovery*, Chapman & Hall/CRC, Boca Raton,

D.E. Sosnowski, D.T. Wilcox (2009), *Basics of Post-Conviction Sex Offender Polygraph Testing*, (in:) D.T. Wilcox (ed.), *The Use of the Polygraph in Assessing, Treating and Supervising Sex Offenders. A Practitioner's Guide*, Wiley–Blackwell, Chichester,

J.G. Speight (2009), *The Scientist or Engineer as an Expert Witness*, CRC Press, Boca Raton,

A. Stein (2005), *Foundations of Evidence Law*, Oxford University Press, Oxford,

F. Taroni, A. Biedermann, P. Garbolino, C.G.G. Aitken (2004), *A General Approach to Bayesian Network for the Interpretation of Evidence*, Forensic Science International, 139, 5–16,

F. Taroni, S. Bozza, C. Aitken (2005), *Decision Analysis in Forensic Science*, Journal of Forensic Science, 50, 4,

F. Taroni (2009), *Bayesian Networks*, (in:) A. Jamieson, A. Moenssens (eds.), *Wiley Encyclopedia of Forensic Science*, Wiley, Chichester,

W.J. Tilstone (2000), *Quality Assurance*, (in:) J.A. Siegel, P.J. Saukko, G.C. Knupfer (eds.), *Encyclopedia of Forensic Science*, Academic Press, San Diego,

W. Twining (2003), *Evidence as a Multi-Disciplinary Subject*, Law, Probability and Risk, 2, 91–107,

W. Twining (2006), *Rethinking Evidence. Exploratory Essays*, Cambridge University Press, Cambridge,

A. Vrij (2008), *Detecting Lies and Deceit. Pitfalls and Opportunities*, John Wiley & Sons, Chichester,

A.K. Webb, M.D. Handler, D.J. Krapohl, J.C. Kircher (2008), *A Comparison of the Objective Scoring System and Probability Analysis*, Polygraph, 37, 4,

F.L. Wettering (2009), *Counterintelligence. The Broken Triad*, (in:) C. Andrew, R.J. Aldrich, W.K. Wark (eds.), *Secret Intelligence. A Reader*, Routledge, London–New York,

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 Identification*, Journal of Forensic Science, 23 (3), 596–601,

J. Widacki (1982), *Analiza przesłanek diagnozowania w badaniach poligraficznych* (*The analysis of diagnostic premises in polygraph examination),* Wydawnictwo Uniwersytetu Śląskiego, Katowice.

Book reviews

EUROPEAN
POLYGRAPH
Volume 4 • 2010 • Number 3 (13)

Review of Materials and Proceedings from Conference on Using the Polygraph in Forensic and Human Resource Studies

Nowadays, polygraph testing is frequently used in most European countries in both criminal cases and recruitment. The amount of cases in which polygraph testing could be applied is increasing. For a practising polygrapher, it is crucial to share experience and problems and ask questions about different situations, because this facilitates problem-solving. A conference on different aspects of polygraph testing for criminal cases and recruitment was held in the Higher Police School in Szczytno, Poland on 12–13 June 2008. For two days experts from Polish universities and the police and invited foreign guests presented historical and current opportunities of polygraph testing including legal regulations. The symposium was addressed to academics, Polish police management, experts testing police officers and professionals from criminal laboratories. The book contains 11 chapters. One of them, presenting a new, computerized polygraph, has the form of a printed slide PowerPoint presentation.

The first chapter, written by Professor Jerzy Konieczny (Andrzej Frycz Modrzewski Krakow University), discusses the preparation of expert opinion in both a criminal and employment setting. The author explains the range of duties and information about cases indispensable for the expert to prepare an examination. Devising an adequate question format is a key issue for a polygrapher. A properly devised question increases the usefulness of the expert opinion and reduces the testing duration. In this chapter Professor Konieczny presents hints on adequate question formulation.

The second chapter presents the history of polygraph use before 1990. The author – Łucjan Wiśniewski, PhD – explains that in Poland the polygraph is also miscalled a variograph.
An invited guest from Lithuania, Vitas Saldžiūnas, considers how they made a transformation of a POT test to EKT (Event Test Knowledge), and obviously the reason for this. This paper, the third chapter, also explains problems even in understanding the definition of “deception” in Lithuanian courts.
The acceptability of using polygraph testing in criminal cases is the main topic of the next chapter. Edward Lewandowski provides an analysis of article 199a of the Criminal Procedure Code, emphasizing requirements for any method to generate reliable proof which can be acceptable in a criminal trial. If those requirements are fulfilled, polygraph testing should be regarded as a reliable method on which expert opinion can be based.
The police officer Piotr Herbowski discusses the usefulness of the POT test and its limitations in certain situations. He also explains why this method is so important for the court by presenting its major advantages, not omitting the considerable risk of missing a “guilty” person.
The next chapter, the fruit of cooperation between R. Kwasiński, M. Tokarski and M. Zubańska, analyzes the influence of polygraph testing on the effectiveness of an investigation.
The authors present their own findings about the frequency of polygraph testing in all Polish districts, the referral source of testing and the type of crime cases in which it was used, including information about the time interval between a crime being committed and the testing process.
A range of opportunities of polygraph use in criminal investigation activities is outlined by Superintendent Marek Abramowicz. He specifies its three main applications: to a suspect, to a witness and to verify criminal version from the investigation. Moreover, the author claims that polygraph testing may be used to control the veracity of an informant and his/her future plans for living and to verify if the intention to get a crown witness status is honest.
It is not only criminals who may be tested. In the following chapter, Superintendent Ewa Reczek defines the legal basis of using psychophysiological testing in the police.
The chapter entitled “Difficulties associated with using polygraph testing in hearing of evidence” contains an interpretation of the Polish Supreme Court’s judgment about the situation when the witness’s statement is the only existing evidence of guilt and it is possible to verify it by testing performed by an expert. In such cases this testing should be undertaken. The author of this paper – Superintendent Piotr Herbowski – claims that it also refers to polygraph testing.
The aforementioned presentation, on a new, computerized polygraph named DIANA 01, is placed in the penultimate chapter. Pictures show software options of the apparatus and its components.

The last chapter, written by Superintendent Iwona Klonowska-Senderska and entitled "Polygraph testing as an element of candidate selection for a chosen group of positions in the police", discusses the legal basis of a specialist course for future polygraphers. The author claims that the course is not well prepared and too short to provide students with even basic knowledge on the subject.

This book can be strongly recommended not only for those who were unable to participate in the symposium, but also for all interested in the opportunities of using polygraph testing in criminal cases and recruitment for the police. The papers described above discuss both theoretical issues and potential difficulties during polygraph testing. Certainly, this is a valuable source of knowledge, which may help to understand the most important aspects of detecting deception from the professional point of view.

Małgorzata Wrońska*

* malgorzata.wronska@yahoo.com.

The Basic Information for Authors

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

Submitted manuscripts must be written in English.

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

Texts for publication should be submitted in the form of normalized printout (1800 characters per page) 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 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: margerita.krasnowolska@kte.pl

Subscription: Terms and Conditions

Krakow Educational Society (Krakowskie Towarzystwo Edukacyjne sp. z o.o.), in its capacity of the distributor of Krakow University (Krakowska Akademia im. Andrzeja Frycza Modrzewskiego) publications, offers the subscription of European Polygraph, quarterly edited by Professor Jan Widacki, LL.D.
Please, send your orders by e-mail to ksiegarnia@kte.pl, including:

- your full name (first and last in case of natural persons; registered business name in case of legal persons),
- address (permanent address or registered seat),
- tax identification number,
- address for delivery of your copies of European Polygraph,
- number of successive issues ordered (minimum 4), and
- number of copies of each issue.

The price of a single copy of European Polygraph is PLN 20 (USD 7.5, € 5.5). Shipment costs will be added on top of the subscription price depending on your country of residence.