A new evaluation of old data regarding the effect of prior expectations on examiners' decisions in the CQT polygraph test; applied vs. pure science perspectives in judging the consequences of the study.

Avital Ginton, Ph.D.

AV-GN Polygraph, Applications & Research, Independent Practice, Tel-Aviv, and Law Faculty, Tel-Aviv University, (part time), Tel-Aviv, Israel

Author Note

The present paper is built upon data from a 1994 experimental study done in Israel National Police (Elaad, Ginton & Ben-Shakhar, 1994) of which the current author was a co-researcher. At the time, the data of the original experiments was collected the author headed all the Polygraph activities in Israel National Police (1977-1994), including Polygraph Field Labs, a Polygraph School, and a Research Unit. In addition, during that period of time, he was also a part-time faculty member in the Department of Psychology at Tel-Aviv University and in the Department of Criminology at Bar-Ilan University, Ramat-Gan, Israel.

Correspondence should be addressed to ginton@zahav.net.il

This is a pre-print version -2017

Abstract

The potentiality of prior expectations of polygraph examiners to affect the outcome of a test is considered a main flaw of polygraph testing. In particular, it has been pointed out that the most commonly used polygraph technique; the Comparison Question Test (CQT) by its very nature is prone to suffer from this weakness. The phenomenon was demonstrated in a study published in 1994 (Elaad, Ginton & Ben-Shakhar, 1994), in which examiners were asked to score blindly tests records after being led to believe that they belonged to examinees who following the tests, had confessed of being guilty, or to examinees that their innocence was proven through confessions made by others. It turned out that when the records were objectively inconclusive, a statistically significant shift of the final scores took place in the direction which was established by the manipulative information that the examiners had received. No effects were found with objectively conclusive records. The current study presents a new analysis of the same data suggesting that when it comes to applications, the derived consequences of these findings are minimal, and the expected impact of prior expectations on real-life polygraph decisions in this regard has been greatly overestimated by polygraph opponents. The new analysis shows that based on the published data, it is estimated that in practical application, the examiners' prior expectations would affect polygraph decisions in only 3% of the total volume of polygraph testing; not really enough to play any significant role against the use of CQT polygraph testing.

Key words: polygraph; CQT; contamination; prior expectations; applied perspective.

Introduction

The use of the polygraph as a means for detecting deception, or more precisely, a technique to find out whether a person is telling the truth or lying about certain facts or information has been a controversial issue for many decades. The controversy is taken place in three different arenas. The practice arena, in which different schools of practitioners are wrestling to win the best-practice cup (e.g. Matte, 1996, Pp.322-465; Krapohl, 2006, 2007; Matte, 2007a, b; Gordon, 2007; Amsel, 1999; Matte & Backster, 2000), the academic arena in which scientific debates are taking place within the ivory towers of the academia (e.g. Ben-Shakhar, Gamer, Iacono, Meijer, and Verschuere, 2015; Elaad, 2015; Ginton, 2015; Ogawa, Masuneoka, and Tsuneoka, 2015; Palmatier and Rovner, 2015a, b; Patnaik and Kircher, 2015; Vrij, 2015) and the third one, perhaps the most popular arena, hosts the heavy loaded differences between the world of polygraph practice and academic perspective (National Research Council, 2003; Ben-Shakhar, 2002; Krapohl & Shaw, 2015; Raskin & Kircher, 2014; American Polygraph Association, 2011; British Psychological Society, 2004). There is room for all three of them, but it seems that when applied science is concerned, the third arena is the most meaningful one.

Most of the controversy has been focused on the Comparison Question Test (CQT), previously known as the Control Question Test (e.g. Vrij, 2008). For the sake of those

who are not familiar with this test, the following is a concise description of its basic structure and procedure.

Briefly, the CQT is administered in the following stages: First, the examiner gets acquainted with the case and a draft of the target questions is prepared. Then, the examiner conducts an extensive non-interrogative pre-test interview with the examinee regarding the relevant issue in which the interviewee is given the opportunity to present his or her version. The examiner also leads the conversation to cover a wider spectrum of attitudes and previous possible misconducts of the examinee in order to lay the foundation for proper Comparison Questions. The series of questions, to be asked later in the actual examination phase of the polygraph test, is finalized during this interaction. The examiner discusses the formulation of the questions with the examinee and ensures that he or she understands them and can give a direct 'yes' or 'no' answer to each question. Then the examiner explains the testing procedure and informs the examinee that the examination is voluntary. The next stage is to attach the examinee to the polygraph, and run an acquaintance short test[1] and following that the actual examination stage starts by asking the prepared series of questions while continuously monitoring the physiological reactions. A typical pre-test interview lasts between 30 to 60 minutes and the test phase about 30-40 minutes.

The questions are mainly of the following three types: (a) Relevant questions - dealing directly with the issue/crime under investigation, such as the 'Did you do it?' type (e.g., 'Did you steal the missing Smartphone from your boss's office last week?'). The number of relevant questions per series is typically 2-4; (b) Comparison questions - focusing on probable past misconducts of the examinee, usually but not necessarily, similar to the issue under investigation. The dynamic of developing and phrasing the

[1] There are versions in which this stage and the previous one are taken place earlier.

comparison questions are designed to elicit a probable lie verbal response from the examinee or at least to cast a doubt in his mind about the truthfulness of his chosen answer. An example of comparison question can be – 'Have you ever stolen anything?' to which the examinee has been maneuvered to respond with a 'NO' answer (e.g. Krapohl and Shaw, 2015; National Research Council, 2003; Raskin and Honts, 2002). The number of Comparison questions is 2-5 per series; (c) Irrelevant questions − focusing on completely neutral issues, (e.g., 'Is today Tuesday?') which are intended to absorb the initial orienting response evoked by the first question, and to enable rest periods between the more loaded questions. The number of Irrelevant questions per series is usually 1-4. Typically, the whole series of question is repeated three to five times in different order with intermittent short breaks between the repetitions.

The decision about deception is derived from comparing the physiological reactions to the Relevant and the Comparison questions. For reasons which are beyond the scope of this article, it is expected (by the CQT proponents) that deceptive ('guilty') examinee will respond to the Relevant questions with more prominent reactions relative to the Comparison ones, whereas the opposite, i.e. more pronounced reactions to the Comparison questions, characterizes a truthful ('innocent') examinee. In case the difference between the questions of the two categories is not clear or does not reach an acceptable degree, the examination is deemed Inconclusive.

The controversy around the CQT covers several issues; one of them is known as the contamination argument. In a critical review of the CQT, Ben-Shakhar (2002) described the essence of this argument as follows 'The CQT contains an element of "contamination". In other words, it is possible that the conclusions made by a CQT examiner are based on information that was in his hands prior to conducting the examination, rather than the physiological measurements recorded by the machine.'

(Ben-Shakhar, 2002, p.118). The danger of having this kind of contamination stems mainly from two sources. First, before the examiner starts the actual testing phase, he gets background information to make sure that he understands the case. Second, although there are rules and standards of how the test should be conducted, its standardization is not totally rigid, and the pre-test interview stage contains flexibility to allow adaptation or a kind of a tailor-made process to fit the specific characteristics of the examinee and the case. While these two elements are considered by the practitioners as a means to improve the accuracy of the CQT, one cannot deny that due to their very nature, they contain the potential for contaminating the test by creating certain expectancies in the examiner mind which may lead to confirmation bias (Kassin, Dror, and Kukucka, 2013). This bias is a well-known and researched phenomenon in cognitive and social psychology that might influence the polygraph test and its interpretation (National Research Council, 2003, Pp. 89-91 & P.138). Yet, one should remember that the existing of a potential for something to occur, doesn't mean that in real life within normal boundaries, it does occur in any meaningful way or at all. This should be left to empirical studies to determine, as was pointed out in the National Research Council report: 'It therefore remains an empirical question whether polygraph test results and interpretations support such hypotheses and whether, in fact, test validity is diminished to any significant degree by examiner or examinee expectancies' (National Research Council, 2003, p.90).

Only two empirical studies from the mid 1990s have addressed this hypothesis (Elaad, Ginton and Ben-Shakhar, 1994, 1998). Both were conducted with polygraph examiners from Israel National Police. The first, aimed to examine the potential effects of prior expectations of the examiners about the guilt of the examinees, on the interpretation of the physiological records. It was found that no effect of prior

expectation could be demonstrated when the records were clear enough to render a decision by examiners with no prior expectations. However, when the records were not that clear to make a call about the veracity of the examinee (Inconclusive records), prior expectations did have a significant effect on the results of the scoring process (Elaad, et al., 1994). Since then, this finding has been the leading empirical scientific demonstration of the danger embedded in the CQT from possible contamination due to prior expectation about the veracity of the examinees (e.g. Vrij, 2008, Ben-Shakhar, 2002). The 1998 study was aimed to examine the possibility that prior expectations of the examiners, could lead to biased outcomes, probably by affecting the manner by which the examinations are conducted. No effect was found in that study. Note that the two studies dealt with event-related type of examination, and the conclusions cannot be generalized automatically to other kinds of polygraph tests, such as pre-employment or periodic tests.

For the intra-academic debate about the scientific nature of the CQT and whether it should be considered an objective scientific test at all, the finding that prior expectations do affect the scoring of the tests (Elaad, et al., 1994) is an important piece of information that undermines the claim about its scientific nature. However, when it comes to applications, a more relevant arena is the one in which the academic world has to face real-life implications and considerations. The present paper is an attempt to function within this arena by relating the 1994's findings to field usage of CQT. Analyzing that data through the prism of an applied perspective become even more important when facing the existing common practice to lean on the findings of the 1994 study as one of the main reasons for disqualifying CQT polygraph testing in the field (e.g. Ben-Shakhar, 2002).

To lay the foundations for this new analysis, a concise description of the original experiments is presented hereinafter.

Two successive experiments with similar paradigms were conducted using records of field CQT polygraph examinations that were drawn from the general pull of criminal polygraph examinations of the Israel Police polygraph lab. In the first experiment, records of fourteen real-life criminal CQT polygraph examinations were given to ten police examiners for blind[2] numerical scoring. Numerical scoring is a semi-objective technique for evaluating the records. The technique was first introduced into the polygraph testing field by Cleve Backster at the late 1950s/early 1960s (see, for instance, Backster, 1963a, b).). The relative strength of reactions to Relevant and Comparison questions is compared along the charts, and a number is assigned to each comparison point to indicate the direction and the degree of the differences between the reaction's strength. A seven points scale ranging from +3 to -3 was used, with plus indicates stronger reactions to the Comparison question, and minus indicates stronger reactions to the Relevant question. Aggregating the numbers along the whole test, results in a final score - the outcome of the test. A primary decision rule is that in order to make a call, the final score should reach a certain cut score, or it is deemed inconclusive. The cut scores used in the study were +/- 6, i.e., the range between +5 to -5 was defined as an inconclusive zone.

All records chosen for that experiment were inconclusive records as was determined by three other examiners who scored them blindly during the selection process. In the second experiment the same ten examiners received for blind scoring twelve conclusive records; six of them indicated clearly deception outcomes and the other six truthful outcomes.

[2] Blind scoring means that the scorer did not know which cases he was scoring or the content of the questions. All he or she got were the physiological records and the kind of questions asked in each segment (e.g. Relevant, Comparison or Irrelevant).

In the first experiment, the 14 records were arbitrarily divided into two sets of seven records each. The two sets were given consecutively to the examiners for blind scoring on their own pace. To manipulate the examiners' expectations, each batch of records was accompanied by different outcome information: In the Guilt-expectation condition the examiners were told that the examinee ultimately confessed of being responsible for the crime, and in the Innocent-expectation condition, they were told that another person had confessed to that crime. Five examiners scored one set under the Guilt-expectation condition while the other five scored the same records under the innocent-expectation condition, and the opposite was done with the other set. The order of the two conditions was counterbalanced across examiners.

Each record was scored 10 times; five times under the Guilt-expectation condition and five times under the Innocent-expectation condition. A within-examiner standardization of the final scores was used for statistical analysis (Ben-Shakhar, 1985). Each final score was transformed into a standard score relative to the mean and the standard deviation of the individual examiner's final scores distribution. The standardized scores given under the two experimental conditions were compared, and the results indicated clearly that prior expectations affect the outcomes of the scoring. The results were statistically significant and reflected quite a strong effect, as it managed to reach the statistical significance with a relatively small sample size[3].

Unlike the results of Experiment 1, in which all records were objectively inconclusive, no effect was found in experiment 2 that used only conclusive records.

Thus, it seems that when the specific physiological information clearly contradicts prior expectations, examiners tend to stick to the physiological information and ignore their prior expectation in making their judgments.

[3]Although not reported in the original paper, a current check for effect size by computing Cohen's d score or r^2 indicates large effect size.

Data analysis from an applied perspective

From academic point of view, these findings are very significant, and it is clear why they have been cited in so many instances to support claims against the use of CQT (e.g. Vrij, 2008; Ben-Shakhar, 2002). However, in order to judge the actual danger of the phenomenon that was demonstrated through the study, one should proceed in the analysis one step further and look at the data vis-à-vis the existing practice of CQT usage.

Table 1 presents the data from Experiment 1 of the original study, however, the final scores given by the examiners are arranged in three categories rather than a continuous scale which was the basis for the statistical analyses mentioned above.

Table1. Frequencies of the classifications made by each examiner under the two experimental conditions in experiment 1.

	Experimental Conditions					
	Guilt expectation			Innocence expectation		
Examiner	NDI	Inconclusive	DI	NDI	Inconclusive	DI
1	0	6	1	4	3	0
2	0	7	0	1	6	0
3	0	7	0	2	5	0
4	1	5	1	1	6	0
5	1	6	0	0	7	0
6	1	6	0	1	6	0
7	0	7	0	1	6	0
8	1	6	0	2	5	0
9	1	5	1	2	5	0
10	1	6	0	2	5	0
Frequencies Across	6	61	3	15	55	0

examiners						
Percentage	8.6%	87.1%	4.3%	21.4%	78.6%	0

NDI = No Deception Indicated, (+6 or higher); DI = Deception Indicated, (-6 or lower) ; Inconclusive zone (+/- 5)

The data presented in the table was published originally in **Elaad, E., Ginton, A., and Ben-Shakhar, G. (1994).** *The effect of prior expectation and outcome knowledge on polygraph examiners' decisions.* **Journal of Behavioral Decision Making, 7, 279-292.** Permission to use it was granted by John Wiley & Sons Ltd.

The continuous scale by its very nature contains more information, and it is much more sensitive to detect differences. However, it does not represent the way the test is used in the field as long as the decision rule in practice is based on cutoff scores separating zones of decisions. In the world of practice, there are only three discrete categories, Deception Indicated (DI); No Deception Indicated (NDI): and Inconclusive. Keeping this in mind, one may question the practical relevancy of the findings which was demonstrated in the continuous scale analysis. What might be relevant are the effects of the prior expectations on the amount or percentage of scores that were maneuvered to cross the cut scores' boundaries both ways.

Remember that objectively (with no manipulated expectation) all the records were Inconclusive, as was determined by three other examiners, namely, the objective scores were within the range of +/-5. Now, taking a look at the distribution under the Innocence expectation, one can see that a shift towards the NDI category which is the predicted direction had occurred. The effect is quite impressive as more than one-fifth of the scores have crossed the cutoff score toward the expected zone. But if we take a look at the distribution under the guilt expectation, we see that 8.6% of the scores have crossed the same cut score that separates the inconclusive zone from the NDI zone. This presumable shift was against the predicted direction, so we might assume that it reflects basic inclination that these examiners had towards NDI scores relative to the three examiners who had scored the records for establishing their inconclusive

nature, and/or random fluctuation. When we now look back at the shift of 21.4% that occurred under the Innocence expectation condition, we might assume that part of it is just the representation of this basic inclination towards the NDI scores and/or random fluctuation. The best estimate we can make about the size of this part is to adopt the figure of 8.6% found under the Guilt expectation condition. Taking this into account, the estimated effect of the Innocence expectation is reduced to 12.8% (21.4% - 8.6% = 12.8%); A smaller effect, but still a considerable one. Here come the results of Experiment 2 and suggest that the actual impact of this contamination on real-life CQT outcomes is much lower, to the point of being almost negligible or at least can be contained quite reasonably as a tolerable flaw.

As mentioned above, Experiment 2 in the 1994 study (Elaad et al., 1994) showed that when the records were objectively conclusive, no effect of prior expectation was found.

The estimated Inconclusive rate in CQT, based upon numerous observations, is revolving around 15% (give or take 5%), and in fact, the American Polygraph Association Standard of Practice requires that any accepted polygraph techniques should have demonstrated objective (blind scoring) inconclusive rate, which does not exceed 20%. (Nelson 2015; American Polygraph Association, 2011; Raskin & Honts, 2002).

All of the above brings to a conclusion that based on the finding of the 1994 study (Elaad et al., 1994) the actual expected impact of prior expectations on field polygraphy, i.e. the percentage of polygraph examinations that their scores are expected to move in any meaningful way and cross the cut scores' boundaries, is the multiplication of 12.8% by at most 20%. i.e. 2.56% for Innocent expectation, and

4.3% by 20%, i.e. 0.86% for Guilt expectation. Of course, these are not precise figures, and one may look, for instance, for confidence intervals, but it certainly gives a fair estimation of the actual impact. In conclusion, the results of the 1994 experiments, although they have clearly demonstrated the unfortunate effect of prior expectations on polygraph scoring, when the real-life impact is concerned the size of these effects is limited to about 3% of the tests.

Discussion

Several factors might have contributed to the fact that the demonstrated effects of prior expectations on scoring the CQT records (Elaad et al., 1994), though being statistically significant, and reflecting quite a large effect size, is rather limited when the actual field implication is concerned. The most important factor in limiting the contamination impact on real-life polygraph usage is the existing of an inconclusive zone which functions as a buffer against the influence of irrelevant factors on the conclusions drawn from the test outcome. The main reason for setting this zone is the need to cope with the adverse effects of random fluctuations or irrelevant random noise (Ginton, 2013), but it also weakens the none-random irrelevant factors such as prior expectations, from playing an undesirable role in affecting the outcome. If it was not for this buffer zone, the prior expectations' effects on real-life decisions would have been much more severe.

Second, taking together the results of the two experiments presented in the 1994 study, have indicated that the prior expectation effects on scoring the physiological activity charts are confined to tests in which the of records were not clear enough and deem Inconclusive. Inconclusive tests resulting from either too small differences between the strength of reactions measured in the Relevant versus the Comparison

questions, and/or a lack of consistency in the directions of the differences found over the several repetitions of the questions. The percentage of inconclusive tests in CQT is between 10%-20% (Nelson 2015; American Polygraph Association, 2011; Raskin & Honts, 2002), thus, a portion of 80% of the tests seems to be "out of the game". The danger does exist for the remaining 20% but assuming the same effect size that was found in the 1994 study its impact on the whole, mass of polygraph examinations is limited to about 3% of the tests' volume, in which prior expectations might move the final scores from the inconclusive zone to conclusive ones, and vice versa, but not from one conclusive zone, to the contrary conclusive zone, to produce opposite conclusions. Had the proportion of Inconclusive records in CQT examinations been larger and/or conclusive examinations were affected too by prior expectations, their expected impact on field polygraphy would have obviously been greater.

The current analysis relies on the asserted differentiation between the prior expectations' effect found among the inconclusive cases and the lack of effect among the conclusive ones. This is based on the findings reported in the 1994 study in which a statistical significant result was demonstrated in a sample of inconclusive cases as opposed to a failure to reach statistical significance with a sample of conclusive records. However, one may raise the point that it does not mean that the measured effects in the two types of cases are indeed different from each other in that regard. Another word, it is not clear whether the vulnerabilities of the two types of cases; inconclusive and conclusive records, to prior expectations effect are significantly different from each other. It is logical to consider this vulnerability to be a continuous variable rather than a dichotomized one, nevertheless, using categories to define these two types of cases as a matter of routine in field polygraphy raises the question are

these two defined categories different significantly from each other with regards to their susceptibility to prior expectations' effect in the common real-life practice. Unfortunately, no attempt was done at the time to get such an indication. The current analysis adds a t-test for repeated measures between the standardized d scores computed in the 1994 study in each of the two samples between the Guilt and the Innocence expectations conditions. The test yields a statistical significant result ($t= 4.23$, $df= 9$, $p<0.01$)[4].

That means that although it is reasonable to consider the prior-expectations-effect variable to be a continuous one that is not affected sharply at any cut scores, the use of the common cut scores defined by practical standards, succeeds in dividing the total volume of polygraph cases into two different qualitative groups; the inconclusive cases which were found to be affected by prior expectations and the conclusive group that as a group seems to be immune to the kind of prior expectations used in the study. That enabled the kind of computation mentioned above, which has led to the conclusion that based on the 1994 findings, in reality, the percentage of polygraph cases that are prone to be affected in any consequential way by the prior expectations' effect of the kind presented in the 1994 study is very low, resides in the neighborhood of three percents of the total volume.

From an academic perspective the importance of the 1994 findings for the ongoing debate about the scientific nature of the CQT and whether it should be considered an objective test at all, still holds even when in practice, the volume of the affected cases is estimated to be very low. But when it is judged from an applied perspective that entails practical considerations one should ask whether the 3% figure, although probably represents thousands of people all over the world, is important enough to

[4] The reason for using t-test for repeated measures is that the same 10 polygraph examiners scored the two samples; the inconclusive and the conclusive ones.

disqualify any CQT polygraph usages. It seems that in order answer this question we should ask ourselves "important for what?"; and forecast the expected consequences, taking into account not only the 3% but also the majority part, namely, the other 97%, and of course to weigh the significance of these evaluations. Yet some people might say that an mistaken result caused by an examiner's bias is very important to the individual, regardless of the frequency of such cases. That is true of course since in the end, any accurate or erroneous outcome concerns individual examinees, affect them, and of course usually very important to them, and therefore, any factor or cause that brings about these outcomes is important, and prior expectations' effect is no exception to that. Examiners and clients who use the outcomes to help them in making their decisions should care to minimize the adverse effects (including prior expectations) and encourage the beneficial ones, and that applies also to society in general. But the cost of any such attitude and act should be evaluated as well, to make sure that the baby is not thrown away with the bath water. Irrespectively to the issue of prior expectations' effects, for every single examinee that unfortunately suffers from false positive (FP) error outcomes, the error is one hundred percent error but for the general society he is but one case that fell into the 10% error rates (Raskin and Kircher, 2014). To put things in perspective, the very fact that unfortunate distinct mistakes do exist should play only a minor role in the general decision-making process of whether to use polygraph or not.

To eliminate possible misunderstanding it should be stressed that the prior expectations' effect bears no clear and direct relation and certainly not a significant one to the issue of test accuracy which the CQT proponents estimate it to be around 90% (e.g. Raskin and Kircher, 2014), and one should not conclude that the whole 3% affected by prior expectations are errors.

Moreover, sometimes we forget or ignore it but prior expectations can be established towards deception or truthfulness, and it is potentially expected that half of the prior expectations might lead to objectively correct outcomes.

Bottom line, from an applied perspective, this is a matter of choice making that should be based mainly on the expected benefits and costs. Polygraph tests are an aid for decision making; polygraph use is a means to improve the quality of the taken decisions. Disqualifying the use of CQT all over the range of its applicability due to the concern that 3% might be critically affected by possible prior expectations will adversely affect the quality of decision making for the rest 97% of the total volume of polygraph testing, which is estimated to be hundreds of thousands per year worldwide.

It should be added that a three percent figure is not something to ignore when polygraph results are used as forensic evidence in criminal trials (Kassin, et al., 2013), but this kind of usage is extremely rare in practice[5], and by no means should this small number be considered a sensible and practical reason to disqualify CQT usage in most real-life other situations.

Two reservations, however, should be made, about this rather small estimated overall impact of prior expectations on scoring the test in CQT field practice. It is built on the results of the 1994 study which any attempt to generalize them to the overall CQT practice, suffers from two weaknesses. First, in practice most examinations are scored and decided by the original examiner who has been involved in the actual running of

[5] The percentage of polygraph cases used as forensic evidence in criminal trials varies from country to country. For example, in Belgium it is quite high and might get to a few tens of percents whereas in USA it is less than one per mille and in Israel, for instance, it is totally forbidden to be used as evidence in criminal trials. However, considering the facts that in USA, the total number of examinations per year is estimated to be half a million and in Israel 25,000 while in Belgium there are only about 500 per year, it is clear that the Belgian impact on the world percentage of polygraph tests used as evidence in criminal courts is very low. The polygraph is used in about 40 countries, but most of them ban its introduction to criminal court as evidence and in those that allow it the total numbers are small.

the test, and his verdict has an impact on the investigation and the future of the examinee. Unlike that, the examiners in the experiments were not really involved with their imposed task of scoring the records. They did not have the sense of responsibility for the examinees' future, nor could they have affected in any way the investigatory process of the specific cases. It is conceivable that when the examiner's psychological investments in a case are bigger, he would rely on his subjective feelings, which are led by his expectations, to a greater degree. The second weakness in generalizing the findings of the study to the overall CQT practice concerns the specific type of examiners who served as subjects in the 1994 study. As mentioned above all of them were examiners from the Israeli National Police. This polygraph unit as a whole was psychologically oriented, with many examiners being psychologists rather than investigators. They belonged to the Scientific Interrogation Laboratory, which was a section in the HQ's Forensic Science Division. This affected their perception about the polygraph role and probably could have resulted in a relatively stronger immunity against irrelevant influence of contaminating factors. This was rather an exceptional group of examiners within the worldwide polygraph arena, meaning that it is not implausible that the common or the average polygraph examiner might be more susceptible to prior expectations influence. At any rate, it is suggested that the Inconclusive Zone plays the most important role in preventing irrelevant influence of the prior expectations on the final calls regardless of the examiner being more scientifically, psychologically or investigatory oriented.

Needless to say that the present analysis and its conclusions are confined to the findings presented in the 1994 study, which times and again have been used throughout the literature not only as a main support for the conceptual arguments against the CQT, but also as demonstrating the danger inherent in CQT usage. Well,

to put things in a proper proportion, as long as no other specific empirical data exist, based on the 1994 data, the empirically demonstrated danger for applied polygraphy is relatively negligible and should not play any significant role in the decision should or should not the CQT been used for detection of deception or more accurately as a means for deduction about deception.

Before making a general closing remark it should be stressed that all the above are related solely to the confirmation bias dealt with in that study, and in principle, other kinds of contamination in polygraph CQT can occur, although not yet been demonstrated empirically (Kassin, et al., 2013).

Following the analysis presented in this paper a general remark which goes beyond any specific case should be made; when it comes to applied psychology it is recommended to assess the actual weight and meaning of any empirical findings on real-life situations by using an applied perspective rather than being satisfied with the one presented by basic or pure science. It might help to keep things in proper proportions.

References

Amsel, T. (1999). Exclusive or Non Exclusive Comparison Question: A comparative Field Study, *Polygraph*, 28, 273-283

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. Special Committee Report. *Polygraph*, 40 (4).

Backster, C. (1963a). Standardize Polygraph Notebook and Technique Guide: Backster Zone Comparison Technique. *Backster School of Lie Detection*, New York, NY.

Backster, C. (1963b). New standards in polygraph chart interpretation – Do the charts speak for themselves? *Law and Order*, 11, 67-68

Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in psychophysiological responsivity. *Psychophysiology*, 22, 292-299.

Ben-Shakhar, G. (2002). A critical review of the control question test (CQT) in: Kleiner, M., (Ed.) *Handbook of Polygraph Testing* (pp.103-126). London, UK: Academic press.

Ben-Shakhar, G., Gamer, M., Iacono,W., Meijer, E., and Verschuere, B. (2015). Preliminary Process Theory does not validate the comparison question test: a comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology* 95, 16-19.

Elaad, E. (2015). Cognitive and emotional aspects of polygraph diagnostic procedures: a comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology* 95, 14-15.

Elaad, E., Ginton, A., and Ben-Shakher, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making.* 7, 279-292.

Elaad, E., Ginton, A., and Ben-Shakher, G. (1998). The role of prior expectations in polygraph examiners decisions. *Psychology, Crime and Law*, 4, 1-16.

Ginton, A. (2013). The importance of the consistency factor in CQT and other polygraph tests. *Polygraph*, 42, 146-162.

Ginton, A. (2015). Good intentions that fail to cope with the main point in CQT: a comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology* 95, 25-28.

Gordon, N.J. (2007). A personal view of Donald Krapohl's validated polygraph techniques. *Polygraph,* 36, 28-29.

Kassin, S.M., Dror, I.E. and Kukucka, J. (2013). The forensic confirmation bais: Problems, perspectives, and proposed solutions. *Journal of applied research in memory and cognition*, 2, 42-52.

Krapohl, D. J. (2006). Validated Polygraph Techniques. *Polygraph*, 35, 149-155.

Krapohl, D. J. (2007). Rejoinder to cretinism by Matte: Closing Comment. *Polygraph*, 36, 50-56.

Krapohl, D.J., and Shaw, P.,K. (2015). *Fundamentals of Polygraph Practice*. Elsevier Inc., Oxford, UK: Academic Press.

Matte, J.,A. (1996), *Forensic Psychophysiology Using The Polygraph: Scientific Truth Verification – Lie Detection*. Williamsville, New York; J.A.M. Publications.

Matte, J. A. (2007a). Critical analysis of Krapohl's validated techniques, *Polygraph*, 36, 29-34.

Matte, J. A. (2007b), A reply to Krapohl's rejoinder to criticism of his article on validated techniques. *Polygraph*, 36, 45-49.

Matte, J.A. and Backster, C. (2000). A critical analysis of amsel's comparative study of the exclusive v. nonexclusive comparison question. *Polygraph,* 29,261-268.

National Research council. (2003). *The Polygraph and Lie Detection, Committee to Review the Scientific evidence on the Polygraph*. Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

Nelson, R. (2015). 2015 update to the APA 2011 meta-analytic survey of validated polygraph techniques. Appendix B in: Krapohl, D.J. & Shaw, P.K. *Fundamentals of Polygraph Practice* (pp. 319-333). Oxford, UK: Academic Press, Elsevier Inc.

Ogawa, T., Masuneoka, I., and Tsuneoka, M. (2015). The comparison question test versus the concealed information test? That was the question in Japan: a comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology* 95, 29-30.

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

Palmatier, J.J., and Rovner, L. (2015b). Rejoinder to commentary on Palmatier and Rovner (2015): Credibility assessment: Preliminary Process Theory, the polygraph process, and construct validity. *International Journal of Psychophysiology* 95, 31-34.

Patnaik, P., and Kircher, J. (2015). Psychophysiological detection of deception and Preliminary Process Theory: a comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology* 95, 22-24.

Raskin, D.,C. and Honts, C.R. (2002) The comparison question test. In: Kleiner, M., (Ed.) *Handbook of Polygraph Testing* (pp. 1-47). London, UK: Academic press.

Raskin, D.C., And Kircher, J.C. (2014). Validity of polygraph techniques and decision methods. In: Raskin, D.C., Honts, C. R., And Kircher, J.C., *Credibility Assessment Scientific Research and Applications* (pp. 65-132). Elsevier Inc. San Diego, CA, USA: Academic Press.

The British Psychological Society (2004), *A review of the current scientific status and fields of application of polygraphic deception detection. Final report* (6 October 2004) from BPS Working Party.

Vrij, A. (2008). *Detecting Lies and Deceit, Pitfalls and Opportunities*. West Sussex, England: Wiley & Sons, Ltd.

Vrij, A. (2015). The protection of innocent suspects: a comment on Palmatier and Rovner (2015). *International Journal of Psychophysiology* 95, 20-21.