Examining different types of comparison questions in a field study of CQT polygraph technique; Theoretical and practical implications

Avital Ginton[1]

AV-GN Polygraph, Applications & Research, Israel

[1] At the time the research was conducted (1981-2), Dr Avital Ginton, headed all the polygraph activities in Israel National Police, including Polygraph Field Labs, a Polygraph School, and a Research Unit. In addition, during that time he was a part-time faculty in the Department of Criminology in Bar-Ilan University, Ramat-Gan, Israel. Lately, he has been affiliated, part time, with the Law Faculty in Tel-Aviv University. Email:ginton@zahav.net.il

Abstract

Two types of questions are most important in the CQT polygraph examination, Relevant questions that deal directly with the case under query, and Comparison questions, which usually deal with past probable misdeeds of the examinees that they choose to deny. The simplistic core reasoning behind the CQT is that the risk of being detected for lying is a threatening situation and like any kind of threat, it automatically triggers the autonomic nervous system to respond with the "fight or flight" type of reaction. For the deceptive examinee, the Relevant questions pose the main threat whereas the truthful examinee, knowing that he is telling the truth on the relevant issue while probably lying to the Comparison questions, perceives the latter as more threatening, considering his goal to appear truthful on the test. Accordingly, the deceptive examinee reactions are focused on the Relevant questions whereas the truthful focuses them on the Comparison ones. Results of the current field study suggest that at least with truthful examinees, Comparison questions, which do not incorporate any lies to be afraid of their exposure, or any lies at all, might function similarly to probable lie questions, by just increasing their salience in a manner that presumably creates some concerns about them. Theoretical and practical implications are discussed in line with the RIG strength notion (Ginton, 2009).

Keywords: polygraph; detection of deception; CQT; field study; Comparison questions; RIG; Relevant Issue Gravity;

Introduction

The measuring of physiological reactions as a means for detecting deception has a long history, emerging in the 20th century in the development of the polygraph and a semi-structured procedure known as the Relevant/Irrelevant test (RIT). In the RIT a series of questions of interest to the investigation - Relevant questions - are asked alongside neutral Irrelevant questions such as "Is today Monday?" (Trovillo, 1939; Matte, 1996). The test is based on the well-known phenomenon that physical or psychological threat elicits a complex set of physiological reactions, mainly in the sympathetic branch of the autonomic nervous system (ANS) that can be measured with a polygraph instrument. Clearly, for deceptive subjects, the relevant questions pose a threat, resulting in an increased physiological activity in the ANS compared to their reactions to the irrelevant questions. This was taken as an indication of deception. Unfortunately, although this method has been found to work well with most deceptive examinees, a considerable portion of truthful examinees show a similar pattern of reactions, resulting in difficulty in separating truthful from deceptive examinees. This was recognized by the polygraph practitioners, and later research has confirmed that the RIT produces a very high rate of False-Positive outcomes (Horowitz, Kircher, Honts & Raskin, 1997; Horvath, 1988).

The need to better separate the truth-tellers from the deceptive subjects eventually brought about a major change in the test, by introducing non-relevant questions that were designed to induce in the innocent suspect a considerable amount of concerns that exceed the concerns he/she might have felt about the relevant questions (Reid, 1947). The differential levels of concern were manifested in differences in the strength of reactions to the two kinds of questions, the relevant vs. the newly introduced type, which were termed Comparison or Control questions (Raskin & Honts, 2002).

Currently, the most commonly used polygraph test is the Comparison-Question-Test (CQT), previously known as Control-Question-Test (Raskin & Kircher, 2014). There are several variations of the test, but generically it consists of the following elements and procedures:

First, the examiner drafts a list of the target questions based on the case material. The examiner then meets the examinee and conducts an extensive non-interrogative pre-test interview in which the examinee is given the opportunity to talk about the offense and present his or her version of the case. During this interaction, the examiner also maneuvers the conversation to cover a wide range 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 stage of the polygraph test, is finalized and 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. It is also mandatory to get a clear agreement from the examinee to be asked each of these questions on the test. The next stage is to attach the examinee to the polygraph, and the actual examination stage takes place by asking the prepared series of questions while continuously monitoring the physiological reactions. 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 laptop from your boss's office last week?"). The number of relevant questions per series is typically 2-4; (b) Comparison questions - focusing on general, non-specific misconducts, usually but not necessarily, similar to the issue under investigation. The dynamic of developing and phrasing the comparison questions are designed to elicit a probable lie 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. Raskin & Honts,2002; NRC, 2003; Krapohl

& Shaw, 2015).   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 question series is repeated three to five times in different order.   A typical pre-test interview lasts between 30 to 90 minutes and the test phase between 30 to 45 minutes.

The decision about deception is derived from comparing the physiological reactions between the Relevant and the Comparison questions.   It is expected that deceptive ("guilty") examinee will react to the Relevant questions with greater strength, whereas stronger reactions to the Comparison questions, indicates a truthful ("innocent") examinee.

A common suggested rationale behind these expectations starts with the well-known phenomenon mentioned above that when an individual is facing a threatening situation or stimuli, the autonomic nervous system is automatically activated as a basic means of protecting the person from being harmed.   In polygraph testing, the questions are perceived as endangering stimuli, which might expose that the subject is not telling the truth, a matter that has some harmful consequences for the examinee.   The deceptive ("guilty") examinee perceives the Relevant questions as posing the greatest threat to his well-being and his/her concerns are focused on these questions.   The Relevant questions are perceived as threatening stimuli also by the truthful ("innocent") examinee, but in the context of being tested for veracity, knowing that he is telling the truth to the Relevant questions while probably lying to the Comparison questions, he focuses his concerns about being detected as a liar on the Comparison questions.   This higher level of concern, driven by the fear of failing the test due to those probable lies, results in increased physiological reactions to the Comparison

questions to a level that exceeds their reactions to the Relevant questions (e.g. Matte, 1996; Raskin & Honts, 2002).

Since the late 1940s when the CQT was first introduced in the field of polygraphy by John Reid (1947), and to a higher degree since the early 1960s when Cleve Backster (1962,1963) introduced his version of it, the differential focus of concern has been adopted by practitioners as the predominant rationale underlying Comparison Question Test (Krapohl & Shaw, 2015). Backster has used the term Psychological Set to conceptualize his main theoretical idea, and this was almost universally accepted by polygraph examiners. By using the term Psychological Set, he meant that the examinee tune in the stimuli that indicate trouble due to the inherent predisposition that any organism holds to detect endangering stimulus, and tune out that which is of a lesser threat for his general well being. In the context of lie detection test the threat comes from the perceived consequences of getting caught lying to any of the presented questions.

Over the years, a few other theoretical frameworks have been suggested by academics as well as practitioners to account for the demonstrated psychophysiological detection of deception in general, and the CQT in particular (Kleiner, 2012; Vrij & Ganis, 2014). . Some have concentrated on explaining why do people react physiologically when they are lying (e.g. Davis, 1961), while other, specifically those which are related to CQT, have focused on explaining the differential strength of reactions that truthful and deceptive examinees are showing to relevant vs. comparison questions. In a recent published book, Krapohl and Shaw (2015) have asserted that " There are currently four ideas that have moved to the fore, and though some may not meet the current understanding of what fully constitutes a "theory", they have at least taken on placeholder status for the purposes of elucidation and testing" (p.201). The first three are relevant to the current study; the aforementioned Backster's Psychological

Set Theory, the Relevant Issue Gravity (RIG) proposed by Avital Ginton (2009), and the Differential Salience suggested by Senter, Weatherman, Krapohl and Horvath (2010).[2] The Psychological Set theory asserts that fear of detection is the operating mechanism that brings about the physiological responding used in polygraphy, and the examinee's Psychological Set determines whether this fear is channeled to the Relevant questions (deceptive examinee) or the Comparison ones (truthful examinee). The Differential Salience theory states that the key factor in CQT is the relative salience of each test question. It represents the subjective importance an examinee gives to a test question as indexed by the physiological response magnitude. It did not rely on the fear of detection though this fear might contribute to salience and by that to affect the size of the physiological response.

The Relevant Issue Gravity (RIG) Strength theory, asserts that the critical point in explaining CQT is the mechanism by which the mirror-like differential strength of reactions that truthful and deceptive examinees are showing to Relevant vs. Comparison questions is developed. The theory states that the relevant issue, by its very nature, is very powerful in attracting and binding the examinee's attention. This is termed Relevant-Issue –Gravity (RIG), and it is postulated that it is stronger for deceptive examinees compared to truthful ones. This in turn, affects the ease by which the examinee's attentional focus can be shifted from the relevant sphere, which is the starting point in any polygraph test, to the comparison sphere and questions. The difference in RIG strength plays a major role in giving rise to two relatively non-overlapping distributions of differences in reaction strength between Relevant and Comparison questions, one for truthful and the other for deceptive examinees. The RIG theory is quite indifferent to the question of why do people react physiologically when they are lying and the proposed alternative explanations, be them fear of detection, relative

---

[2]The fourth one, offered by Palmetier and Rovner (2015) is more focused on the theoretical framework that accounts for the kinds of physiological reactions shown in polygraph examination. In their article, they have suggested that the Preliminary Process Theory (PPT) proposed by Robert Barry (1990) in a totally different context can be adopted and adapted to explaining polygraphy.

salience or any other suggested reason for that matter. Instead, it concentrates on the notion

that the examiner uses baits to attract the examinee's attention and divert it from the relevant

sphere to the comparison one. In principle the bait can take any kinds of form and the ease by

which the shift of attention occurred is negatively correlates with RIG strength which in turn

correlates with deception.

Traditionally, the Comparison questions deal with probable past misconducts or misdeeds that

bear a close resemblance to those which are under the current investigation. However,

occasionally, the examinee refuses to answer these questions, saying that he/she did not come

to take a test on those questions and insists that the questions will be strictly limited to the

relevant case[3]. There are several common practical ways to overcome such situation and stay

within the probable lie arena. Yet,one should remember that getting probable lie answers is

not a purpose on its own. Rather, the development of the CQT and the Comparison questions

mainly aimed to eliminate the innocent examinees from the "deception indicated" category,

i.e. to reduce the RIT False Positive outcomes. Asking probable lie questions seems to do the

work to a great deal by inducing differential levels of concerns in truthful versus deceptive

subjects (Ginton, 2012; Raskin & Honts,2002; Raskin & Kircher, 2014), In a sense, it seems

logical that concern can be developed in the mind of the examinee by the fear that his or her

lies might be detected in the test, particularly because the test is defined as a lie detection

test[4]. However, there are other ways by which concerns can be established. The Directed-Lie

(DL) technique is an example to that. In this technique, the examinees are instructed to lie

about some past petty misconducts of them, by giving a "No" answer to questions such as

"During the first 25 years of your life, did you ever tell even one lie? or "Did you ever commit

a minor traffic violation?" (Handler & Nelson, 2009; Krapohl & Shaw, 2015; Raskin &

---

[3] Experience shows that the rate of these incidences is higher among "professional criminals".
[4] Some scholars, do not concur with this, claiming that it is illogical to expect an innocent examinee to concern more about the Comparison questions, because he realizes that the main issue is covered by the relevant and not the comparison questions (e.g. Ben- Shakhar, 2002)

Honts, 2002). The examinees are told that the role of these questions is to show how they are reacting when lying, and by that to enable seeing whether their reactions to the case questions are the same or different compared to the lying questions. They are also instructed to think of a particular time when they did lie in the past and to have that in their mind when answering these questions every time along the test, since it helps to confirm that they are responding appropriately throughout the test and continue to be suitable subjects. It is assumed that the truthful examinees will focus on showing that they are suitable subjects and on clearly demonstrating that their reactions when lying are different from their reactions to the Relevant questions and by that divert their concerns from the Relevant questions to the Comparison ones (Raskin & Honts, 2002). Results from both laboratory (Horowitz, Kircher, Honts & Raskin, 1997; Honts and Reavy, 2015) and field studies (Honts and Raskin, 1988) are consistent in showing no significant differences in accuracy from the probable-lie method. The success of the Directed Lie Comparison questions (DLC), is incompatible with Backster's Psychological Set theory as clearly no fear of detection is involved in reacting to them. Conversely, it is compatible with both the Differential Salience and the RIG accounts.

By the very fact that the Directed-Lie technique is built upon telling lies, it seems to have face validity and makes it easy to administer. However, practically as well as theoretically, it is important to find out whether Comparison questions that do not have clear face validity, and are unrelated to lies can also function as a means to eliminate the innocent examinees from the "deception indicated" category, which is a must in the process of differentiating between deceptive ("guilty") and truthful ("innocent") examinees. As for now, no empirical study addressing this issue has been published. The following presents a field case study that was

conducted internally in the Israeli Police Polygraph unit in the early 1980s, and never been published, which gives a partial answer to this question[5].

Method

Procedure and Material

In the early 1980s, the Israel Police Polygraph Lab was involved for almost a year in a serious high profile investigation of an ongoing serial arson case that occurred in a main Air-Force base. Almost 300 suspects, most of them soldiers who served in that base, were polygraphed throughout the investigation, using the CQT method[6]. Due to certain circumstances which characterized the dynamic of that investigation, polygraph examinations of a small subgroup of suspects were used for a unique field study in which the relative effectiveness of three types of Comparison Questions (at that time, the common term was Control Questions) were evaluated.

The test was a generic CQT[7] with three relevant questions;

1) Primary involvement – "Did you, in person, set fire to one or more warehouses in the base?"

2) Secondary involvement – "Were you involved in setting fire to one or more warehouses in the base?"

3) Knowledge – "Do you know for sure, who set fire to one or more warehouses in the base?"

---

[5] The study was considered to be an internal research of the Scientific Interrogation Lab of Israel national police and never been published officially outside the Lab. The essence of the research was first introduced to the polygraph community in 2007, embedded in an oral presentation (Ginton, 2007) and was described in a published paper in 2009 (Ginton, 2009); however, only after digging into archive's material it became possible to report the study in more details as it is done here, and present it within the current theoretical and practical perspective.

[6] Some of them had additional series of Concealed-Information-Test also known as Guilty Knowledge-Test (Ben-Shakhar & Elaad, 2002)

[7] The specific kind of CQT format used was the Israel Police Modular Test (IPMT), which is a flexible generic version relative to the formal rigid kinds of formats commonly used in most other places (Ginton & Ber, 1992)

Three different types of questions were used for comparison; one regular probable lie question, commonly applied in arson cases, which was related to past deeds of the examinees; one hypothetical question regarding a potential future misdeed of the examinees, and one bizarre question. On top of that, the "Knowledge" relevant question was always preceded by a "Suspect" question – "Do you suspect any person of being involved in the arsons?", which was used for comparison with the "Knowledge" question solely.

The common probable lie question was - "Have you ever damaged any public property?"   In case the examinee admitted of doing such a thing, he was asked to explain, and then the question was rephrased "Other than what you have told me, have you ever……..?"

The hypothetical question was - "Assuming you have very good reasons and opportunity, would you damage any public property out of rage or for pure fun?"   All the subjects gave a "NO" answer to this question.

The bizarre question was - "Do you like blue in particular?"

Before the "color," question was introduced to the examinee, the following introductory remark was made:

"Look Ron, I am about to ask you a question that might look a kind of bizarre to you. But believe me, it is a very important question, otherwise I would not waste my time asking it. So, think very carefully before you answer me, – Do you like the color of blue in particular?"

Now, regardless of the examinee's answer, the interview went on to discuss the "blue" issue for another few minutes, relating to his/her habits and personality traits and by so doing increased the salience of this question. In case the examinee answered that blue is his favorite color the examiner asked him whether he considered his attraction to blue to be abnormal or pathological in its nature.

The final phrasing of the question was in accordance with this conversation, aiming to get a "NO" answer. Thus, either it was "Do you like the color of blue in particular?" or "Do you

consider your attraction to blue to be extremely abnormal?" and eventually, the chosen answer by the examinees were always "NO".

Decisions for the criminal investigation were taken by comparing the Relevant questions only to the "regular" probable lie and the "suspect" Comparison questions.

The other two Comparison questions, the Hypothetical and the Bizarre, were added only as part of the study. The study aimed to compare the three kind of Comparison questions ("suspect" question excluded) to see whether there are any differences between them with regards to the strength of physiological reactions that they induced in the examinees.

There were 21 examinees that went through this procedure[8]. Based on the comparison between the Relevant questions to the regular Comparison questions, most of them (16) were found NDI (No Deception Indicated) one was found DI (Deception Indicated)[9] and four INC (Inconclusive), and were scheduled to re-exam on another day. The ground truth has been established later by the fact that the case was solved when the actual culprit was caught (not by polygraph), gave a full confession and also supplied some more evidence. Apparently, it was clear that all 21 subjects were indeed innocents[10].

All tests were conducted with Lafayette field analog polygraph instruments, which measured Respiration, Cardiovascular changes and Electrodermal activity.

The whole series of questions was repeated 3-5 times with short breaks between repetitions as it is usually done in CQT procedure.

Comparing the three types of Comparison Questions

---

[8] Twelve were tested by the author of this paper and 9 by other examiners.

[9] When the study was first mentioned in writing (Ginton, 2009), the one DI outcome was omitted from the description by mistake.

[10] Originally, in the 1980s, the relevant questions in each examination were compared also to the Hypothetical and the Bizarre comparison questions, and it was found that the outcomes were very similar to those which were established by comparing them to the conventional Probable Lie Comparison question (Ginton,2009). But unfortunately, the exact data to enable statistical analysis of this could not be traced.

As part of the study, in every repetition the physiological reactions of the three types of Comparison questions were compared to each other (the "suspect" question was excluded) and a rank order of the reactions' relative strength was established by an overall clinical judgment. The question judged as producing the strongest reactions, received 1 point, the second 2 points and the weakest being in the third position received 3 points. In case of similar strengths, a tie rank order was used. Thus if all three questions showed the same strength of reactions, each one received 2 points; a tie of two, resulted in either two scores of 1.5 and the third question got a score of 3 or two scores of 2.5 and 1 point to the third question. The judgment was done by looking at the charts, on which the comparison questions were identified by numbers that indicated their belonging to the category of comparison question. The 21 examinations were judged by three examiners separately. One of them knew the meaning of the three Comparison questions while the other two were blind as to the fact that two of the questions were not the usual probable lie questions. A total score per question given by a judge in each examination was established by summing up the scores across all the repetitions. Those totals served to indicate the judge's evaluation of the relative strength of the three questions compared to each other and were used for ranking them in order. The ranks given by the three judges were averaged to set one rank order, per each examination. A one-way ANOVA for repeated measures was administered to test the differences between the strength of reactions induced by the three types of questions.

Results

The ranks given by the three judges are presented in table 1. It is clear from this data that the rank orders made by the judges were quite similar, and the inter-correlations presented in table 2 indicate that.

Table 1: Rank order of reactions strength, measured in three different types of comparison questions, in 21 individual CQT examinations; Ranked by three judges.*

*The strongest reaction gets1, the weakest gets 3. See text for further clarifications

Judge Ss	Regular Probable lie					Hypothetical					Bizarre				
	I	II	III	Mean	Rank of Means	I	II	III	Mean	Rank of Means	I	II	III	Mean	Rank of Means
1	2	2	2	2	**2**	3	3	3	3	**3**	1	1	1	1	**1**
2	1	1	1	1	**1**	3	3	3	3	**3**	2	2	2	2	**2**
3	1	1	1	1	**1**	3	3	2	2.67	**3**	2	2	3	2.33	**2**
4	2	2	2	2	**2**	2	2	2	2	**2**	2	2	2	2	**2**
5	1	1	1.5	1.17	**1**	2	3	1.5	2.17	**2**	3	2	3	2.67	**3**
6	2	2	1	1.67	**2**	3	3	3	3	**3**	1	1	2	1.33	**1**
7	2.5	2.5	2.5	2.5	**2.5**	2.5	2.5	2.5	2.5	**2.5**	1	1	1	1	**1**
8	1	1	2	1.33	**1**	2	2	1	1.67	**2**	3	3	3	3	**3**
9	3	3	3	3	**3**	1.5	1.5	1.5	1.5	**1.5**	1.5	1.5	1.5	1.5	**1.5**
10	1	2	1	1.33	**1**	2	1	2	1.67	**2**	3	3	3	3	**3**
11	1.5	1.5	1.5	1.5	**1.5**	1.5	1.5	1.5	1.5	**1.5**	3	3	3	3	**3**
12	1.5	2	1	1.5	**1.5**	3	3	3	3	**3**	1.5	1	2	1.5	**1.5**
13	3	3	2	2.67	**3**	1	1	1	1	**1**	2	2	3	2.33	**2**
14	3	2	2.5	2.5	**2.5**	2	3	2.5	2.5	**2.5**	1	1	1	1	**1**
15	1.5	1.5	1.5	1.5	**1.5**	3	3	3	3	**3**	1.5	1.5	1.5	1.5	**1.5**
16	3	3	3	3	**3**	2	2	2	2	**2**	1	1	1	1	**1**
17	2.5	2.5	2.5	2.5	**2.5**	1	1	1	1	**1**	2.5	2.5	2.5	2.5	**2.5**
18	1	1	1	1	**1**	2	2	2	2	**2**	3	3	3	3	**3**
19	3	3	3	3	**3**	1	1	1	1	**1**	2	2	2	2	**2**
20	3	3	3	3	**3**	2	2	1.5	1.83	**2**	1	1	1.5	1.17	**1**
21	2	2	1	1.67	**2**	3	3	3	3	**3**	1	1	2	1.33	**1**
Total			M=1.94		**41 Ranked 2nd**			M=2.14		**46 Ranked 3rd**			M=1.91		**39 Ranked 1st**

Table 2: Correlations (Pearson) between rank-order scores given by three Judges in three different types of comparison questions

		Judge I	Judge II
REGULAR PROBABLE LIE QUESTION	Judge II	0.911*	
	Judge III	0.825*	0.737*
HYPOTHETICAL COMPARISON QUESTION	Judge II	0.877*	
	Judge III	0.888*	0.788*
BIZARRE COMPARISON QUESTION	Judge II	0.953*	
	Judge III	0.864*	0.820*

* $P<0.001$; df =19

A one way within-subjects ANOVA (repeated measures) of mean rank orders given by the three judges is presented in Table 3.

Table 3: One-way ANOVA with repeated measures of mean rank orders given by the three judges, between three types of comparison questions

Summary Values

Values	REGULAR	HYPOTHETICAL	BIZARRE		Total
n	21	21	21		63
sum	40.84	45.01	40.16		126.01
mean	1.9448	2.1433	1.9124		2.0002
sumsq	90.1134	106.7645	87.8934		284.7713
SS	10.6893	10.2931	11.0922		32.7313
variance	0.5345	0.5147	0.5546		0.5279
st. dev.	0.7311	0.7174	0.7447		0.7266

Variances and standard deviations are calculated
with denominator = n-1.

ANOVA SUMMARY and EFFECT SIZE

Source	SS	df	MS	F	P
Treatment	0.6567	2	0.328	0.41*	0.666407
Ss	0	20			
Error	32.0745	40	0.802		
Total	32.7313	62			
Effect size Partial eta-squared**	0.02006***				

* Not significant.
** The one-way
repeated-measures ANOVA. RUTGERS – The state university of New-Jersey, Spring 2013
*** The very small figure is considered as indication of no-effect. http://www.psychometrica.de/effect_size.html

The difference between the three questions is far from being statistically significant (F (2,40) = 0.41, P= 0.666) and the very small value of F indicates that on ordinal scale the relative strength of the physiological reactions in quite similar for all three questions.   Computation of Effect Size of the difference between the three questions using Partial eta-squared, gave such a low effect size (0.02006) that it is considered to indicate No-Effect at all (Lenhard & Lenhard, 2016).

Discussion

Based on the traditional probable lie questions for comparison (PLC), the accuracy rate of the

decisions in this sample (excluding four inconclusive/no-opinion outcomes), was 0.94 with

0.06 False Positive. These figures are a bit higher than the mean accuracy rates reported in

the last two decades times and again for single-issue CQT tests but still fall within the 95%

confidence interval   (American Polygraph Association Ad-Hoc Committee Report on

Validated Techniques, 2011; Ginton, 2012; Nelson, 2015; NRC, 2003; Raskin & Kircher,

2014). Besides the traditional PLC that usually deals with past probable misbehavior of the

examinee, the examinees in the present study were exposed also to two different kinds of

comparison questions: A hypothetical question regarding a potential future misdeed of the

examinee, and a bizarre question that has nothing to do with misbehavior and therefore, no deceptive answer was expected. This bizarre question was introduced in a manner that increased its salience and probably induced some concern about its role in the test. Comparing the three kinds of questions on an ordinal scale with regards to the relative strength of physiological reactions elicited by them show that no difference between them could be found. That can be taken as an indirect indication that their effect on truthful examinees in CQT is similar, which means that in this sense, they can function properly as comparison questions.

In principle, any variable that its values distribute differently in truthful population compared to deceptive population with little or no overlapping between these two distributions can serve as a platform for making a probabilistic judgment about deception. All which is needed aside from identifying the two distributions is having a proper yardstick to measure the specific value that the examinee holds on that variable. The variable that is used by the Comparison Question Test (CQT) for differentiating between liars, and truthful examinees is the difference in the strength of physiological reactions between Relevant and Comparison questions. By relying on the variance of this variable, the CQT accuracy rate of decisions in event-specific cases is found to be in the region of 90% (excluding inconclusive results, i.e. no opinion tests) (American Polygraph Association Ad-Hoc Committee Report on Validated Techniques, 2011; Ginton, 2012; Nelson, 2015; NRC, 2003; Raskin & Kircher, 2014).

Other than empirical knowledge, what is the sense to expect this variable to provide two relatively non-overlapping distributions, one for the truth-tellers and the other for the deceptive examinees? In other words, what is or are the theoretical rationales that underlie the Comparison Question Test?

Three alternative theoretical explanations were mentioned in the introduction, of them the Psychological Set theory proposed by Cleve Backster in the early 1960s has been the most

popular theory within the polygraph practitioners' community for almost fifty years. Fundamentally, it is built upon three bricks: A permanent predisposition to react to any threatening stimulus or situation, by activating the autonomic nervous system (ANS) in a pattern known as the "fight or flight" reaction is the first brick in the theory. Backster used the concept "psychological set" to term this predisposition and was the first to introduce it to the polygraph field (Matte, 1996). The second brick was the notion that when a person is confronted with two or more threats in the same time frame, he would focus his reactions on the most threatening one at the expense of relatively ignore the others. Backster has termed this phenomenon, the anti-climax dampening effect. Then, comes the third brick which states that probable-lie comparison questions, perceived by the truthful examinee to be more threatening to his self-preservation compared to the relevant questions, while the deceptive examinee perceives the relevant ones to be more dangerous to him (Matte & Grove, 2001). These three bricks comprise in the eyes of most polygraph practitioners, the theoretical rationale that underlies the Comparison Question Test, or in a more scientific oriented rephrasing, the rationale for expecting the relatively non-overlapping nature of the two distributions. Technically, that is done by presenting comparison questions that deal with past behavior of the examinee, focusing on potentially misconducts or misdeeds, and maneuver the examinee to lie about them or at least not to be sure whether his chosen answer is true. Another variation of probable lie comparison question, though banned by Backster specific technique, does not mention potential past misconducts of the examinee rather it deals with more general aspects, which are formulated in questions like "Are you really an honest person?". However, the examiner leads the conversation to make it clear that past behavior is a main factor in determining whether the examinee is really an honest person. What is threatening about the comparison questions in the eyes of the truthful examinee is the possibility that he might be caught and tagged as a liar, although he was telling the truth about

the relevant issue, and that might have some undesirable consequences. Backster stated that the "psychological set" of the truthful examinee is tuned on the Comparison questions whereas the deceptive examinee's "psychological set" is tuned on the Relevant questions. Backster insisted on the principle that the "Psychological set" of the examinee cannot be tuned to focus in the same time on both types of questions. At any rate, unless one may assume unconscious mechanism, it seems that paying attention to a question is a prerequisite for developing concern about it and reacting physiologically accordingly. As mentioned above, a key point in this theory is that the truthful examinee is afraid of being caught as a liar because he gave deceptive answers to the Comparison questions. Yet, there are indications from laboratory research that fear of detection did not appear to be a necessary condition for CQT to work pretty well with probable or directed lie comparison questions (Krapohl & Shaw, 2015, p.203). Some evidence to this could be traced also in one field validity study with Directed Lie comparison questions (Honts & Raskin, 1988). These findings of course have challenged the Psychological Set theory and the results of the present field study add to the challenge by suggesting that there is no need to build comparison questions on past misdeeds of the examinee. Nor is there any need to build them on any kinds of lies at all, be them spontaneous or directed.

Lying on the comparison questions doesn't seem to be a necessary condition for getting the CQT to work properly at least in removing the truth-teller from the "deception indicated category." It appears that increasing the salience of a question as was done with the "blue" question in a manner that probably created some concerns about it, was enough to induce the same strength of reactions that were detected with a probable lie question. Note that although salience and concern are correlated, they are separate variables and in the present study, no effort was made to differentiate between their effects. Thus, it is not clear whether a mere increase in salience is enough for producing that outcome, or the concern that could have been

developed in the examinees about the meaning of this question was a necessary factor together with the elevated salience of it. The results of the present study are compatible with the notion of Differential Salience presented by several polygraph researchers (Handler & Nelson, 2007; Krapohl, 2001; Senter et al., 2010) and the Relevant Issue Gravity theory that was proposed by Avital Ginton in 2009 (Ginton, 2009; Krapohl & Shaw, 2015). The Differential Salience proposition suggests that "the relative salience of each test question was the operating mechanism that brings about the physiological responding used in polygraphy" (Krapohl & Shaw, 2015, p.203), but is mute about the mechanism by which a reversed differential salience and focus are created in the truthful examinee compared to the deceptive one. Also, although not mentioned specifically, based on the published material it seems that the proposed theory assumes that probable or directed lying is a necessary factor for achieving the kind of salience, which is needed for the CQT to function. Unlike that, the Relevant Issue Gravity theory proposes a mechanism that explains how this reversed pattern originates from a variable that behaves differently among the truth-tellers and the deceptive examinees. The theory states that a main factor in giving rise to two relatively non-overlapping distributions, one for truth-tellers and the other for deceptive examinees, is the difficulties that the deceptive examinee has in diverting his attention from the relevant sphere to the comparison ones. Unlike the deceptive examinee, it is much easier to divert the attention of the truthful examinee towards the comparison questions and by that to increase the probability that his concerns will focus on them. The difficulty in diverting attention from the relevant to the comparison sphere stems from what is termed in the theory, the Relevant-Issue-Gravity (RIG). This concept is defined as the compelling force arising from an aggregation of qualities that the relevant issue possesses and which interact with circumstantial and personal factors to capture and bind the examinee's attention.

It is postulated that the RIG strength is greater for guilty examinees than it is for innocents due to several characteristics that distinguish the state of mind of the guilty person. An example might be the autobiographical episodic memory traces that in most cases the guilty person, unlike the innocent, carries from being involved with the relevant event under investigation. Those memory traces make it harder for him to divert his/her attention from the relevant to the comparison questions once the comparison issues are inserted by the examiner into the pretest interview. The ease with which the attention is diverted correlates with the relative strength of reaction measured in the Relevant vs. Comparison questions. Thus if it is found, for instance, that the strength of reactions to the comparison questions is greater than to the relevant, one can take it as an indication that it was relatively easy to divert the examinee's attention towards the comparison questions, which means that the RIG strength was relatively weak and therefore, the examinee probably an innocent subject.

This is of course not an all-or-none phenomenon. Attention and reactivity shift from the relevant to the comparison sphere and questions as a continuum in inverse proportion to the RIG strength. The difference in RIG strength between the truthful and deceptive examinees is manifested by having two relatively non-overlapping distributions of RIG strength, which affect the two distributions of differences in reaction strength that are used for making decisions about deception. That is to say, that RIG theory suggests that the differential distribution of the difference of reaction strength is a derivative of a differential distribution of RIG strength.

A major way to evaluate RIG strength is to estimate or measure the amount of effort that is needed for diverting the attention away from the relevant issue and questions, and/or the degree of success in accomplishing this shift of attention.

In the context of CQT, this is done by presenting verbal baits to capture the examinee's attention and shift it from the relevant to the comparison sphere through the dynamic that

characterizes the process of constructing the comparison questions and the final phrasing of them. Maneuvering the examinee to lie about past misbehavior as it is done with the "classic" probable lie CQT, is one way to divert the attention, but in accordance with the RIG strength theory by no means the only one. More than that, a clear derivative of the theory is that, in principle, any attention attracting means, including such that has no connection what so ever with any kind of lying, can function as a comparison question. Results of the current field study support this notion by demonstrating that on an ordinal scale, the three types of Comparison questions produced the same strength of reactions. The RIG theory also explains the success of the Directed Lie approach in which the instructions and the explanations given to examinees concerning the Directed Lies, function as baits for attracting their attention and diverting it from the Relevant to the Comparison questions. More recently the concept of cognitive load has been adopted as a core factor in explaining the demonstrated differential strength of reactions between Relevant and Comparison questions, in deceptive vs. truthful examinees (Honts, 2014; Honts & Reavy, 2015). It is claimed that elevated cognitive loads relating to certain questions results in increased physiological reactions to those questions. This is compatible with the differential RIG strength hypothesis since the cognitive   load is considered a contributing factor in establishing the actual specific strength of the RIG in each individual case. In addition, cognitive/mental load affects the quality and the degree of attraction that the "baits" for diverting attention from the relevant to the comparison sphere will have. In the current study, for instance, the way the bizarre "blue" question was introduced presumably generated a lot of mental processing in the examinee in trying to figure its meaning to the test, along with digging in his or her own semantic and long-term memory, which increased the mental load and mental efforts invested in coping with this question (e.g. De Jong, 2010; Croizet, J.C., Despres, G., Gauzins, M.E., Huguet, P., Leyens, J.P., and Meot, A., 2004). That in turn, resulted in elevated physiological reactions.

Regarding the practical implications of the study, unless it gets support from further research it is not suggested to exchange the "classic" Comparison questions for Bizarre or Hypothetical kinds of questions as a default. However, in case of difficulties in the use the regular types of questions, these could provide an alternative, although the outcomes must be treated with caution.

Finally, due to the specific circumstances, only tests of truthful examinees were analyzed in the current study. Obviously, this is a weakness of the study as the results cannot be generalized automatically to deceptive subjects and in fact no separation of truthful from deceptive examinees has been demonstrated. Therefore the drawn conclusions should be taken only as a partial rather than an overall support of the RIG strength theory. However, keeping in mind that the use of comparison questions is mainly aimed to remove truthful examinees from the "deception indicated" category of the RIT (see above), their effect on truthful examinees is more critical from both theoretical and practical perspectives. This, plus the rarity of field data of the kind that is presented in the current study, brings about the conclusion that it seems right to share the study with others, regardless of its inherent weaknesses.

References

American Polygraph Association Ad-Hoc Committee Report on Validated Techniques, (2011*).* Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph* 40 (4) .

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

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

Barry, R.J., (1990). Preliminary process theory; towards an integrated account of the psychophysiology of cognitive processes. *Acta Neurobiologiae Experimentalis*, 56, 469-484.

Ben-Shakhar, G., (2002). A critical review of the control questions test (CQT). In: Kleiner, M., (Ed.). *Handbook of polygraph testing*. (Pp. 103-126). San Diego, CA; Academic Press.

Ben-Shakhar, G., and Elaad, E., (2002). The guilty knowledge test (GKT) as an application of psychophysiology: future prospects and obsticles. In: Kleiner, M., (Ed.). *Handbook of polygraph testing.* (Pp. 87-102). San Diego, CA; Academic Press.

Croizet, J.C., Despres, G., Gauzins, M.E., Huguet, P., Leyens, J.P., and Meot, A., (2004). Stereotype Threat Undermines Intellectual Performance by Triggering a Disruptive Mental Load, *Pers Soc Psychol Bull* 2004; 30; 721, DOI: 10.1177/0146167204263961

Davis, R.C., (1961). Physiological responses as a means of evaluating information. In: A.D. Biderman and Zimmer (eds) The Manipulation of Human Behavior. (Pp. 142-168). New York; Wiley.

De Jong, T., (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Sciences*, 38,105-134; DOI: 10.1007/s1 1251-009-9110-0.

Ginton, A., (2007). Reframing the notion of psychological set and the role of attention in CQT polygraph examinations. *American Polygraph Association 42nd annual seminar,* New Orleans, August 2007.

Ginton, A., (2009). Relevant Issue Gravity (RIG) strength – a new concept in PDD that reframes the notion of psychological set and the role of attention in CQT polygraph examinations. *Polygraph* 38 (3) 204-217.

Ginton, A., (2012). A non-standard method for estimating accuracy of lie detection techniques demonstrated on a self-validating set of field polygraph examinations. *Psychology, Crime and Law*, DOI:10.1080/1068316x. 2012. 656118; (2013), v.19, 577-594

Ginton, A., & Ber, Y. (1992). *The polygraph doctrine – Understanding polygraphy –Chapters in theory and practice.* Unpublished Internal Manuscript (in Hebrew). Scientific Interrogation Lab, Israel Police.

Handler, M., and Nelson, R., (2007). Polygraph terms for the 21st century. *Polygraph* 36 (3), 157-164.

Handler, M., and Nelson, R., (2009). Utah approach to comparison question polygraph testing. *Polygraph*, 38 (1) 15-33.

Honts, C.R., and Raskin, D.C., (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16, 56-61.

Honts, C.R. and Reavy, R., (2015). The comparison question polygraph test: A contrast of methods and scoring. *Physiology and Behavior*, 143, 15-26.

Horowitz, S.W., Kircher ,J.C., Honts, C.R. and Raskin, D.C., (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108-115.

Horvath, F.S., (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration*, 16, 198-209.

Kleiner, M., (2002). Physiological detection of deception in psychological perspectives: A theoretical proposal. In: Kleiner, M., (Ed.), *Handbook of polygraph testing.* (Pp. 127-182). San Diego, CA; Academic Press.

Krapohl, D. J., (2001). A brief rejoinder to Matte & Groves regarding "psychological set". *Polygraph* 30 (3), 203-205.

Krapohl, D.J. and Shaw, P.K., (2015). *Fundamentals of polygraph practice* San Diego, CA; Academic Press/ Elsevier

Lenhard, W. & Lenhard, A. (2016). *Calculation of Effect Sizes*. available: https://www.psychometrica.de/effect_size.html. Bibergau (Germany): Psychometrica. DOI: 10.13140/RG.2.1.3478.4245

Matte, J.A., (1996). *Forensic Psychophysiology using the Polygraph*, Williamsville, NY; J.A.M. Publications.

Matte, J.A. and Grove, R.N., (2001). Psychological set; its origin, theory and application. *Polygraph* 30 (3), 196-202.

NRC - National Research Council (2003). *The Polygraph and lie Detection*. 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. and Shaw, P.K., (2015). *Fundamentals of polygraph practice*. (Pp. 319-333). San Diego, CA; Academic Press, Elsevier.

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

Raskin, D.C. and Honts, C.R., (2002). The comparison question test, In: Kleiner, M., (Ed.). *Handbook of polygraph testing* (Pp. 1-48). San Diego, CA; Academic Press.

Raskin, D.C. and Kircher, J.C., (2014). Validity of Polygraph Techniques. In:Raskin, D.C., Honts, C.R. and Kircher, J.C., (Eds.) *Credibility Assessment; Scientific Research and Applications*.(Pp. 65-132)*.* San Diego, CA; Academic Press Elsevier.

Reid, J.E., (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology,* 37 (6)542-547.

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

Trovillo, P.V.(1939). A history of lie detection. *Journal of Criminal Law and Criminology, 29* (6),848-881.

Vrij, A., and Ganis, G., (2014). Theories in Deception and Lie Detection. In: Raskin, D.C., Honts, C.R. and Kircher, J.C., (Eds.) *Credibility Assessment; Scientific Research and Applications.* (303-374). San Diego, CA; Academic Press, Elsevier.