Journal of Behavioral Decision Making, Vol. 7, 279–292 (1994)

The Effects of Prior Expectations and Outcome Knowledge on Polygraph Examiners' Decisions

EITAN ELAAD AND AVITAL GINTON
Division of Identification and Forensic Science, Israel National Police Headquarters

GERSHON BEN-SHAKHAR
The Hebrew University of Jerusalem, Israel

ABSTRACT

The present study deals with the question of whether judgments made by experts working in familiar contexts are affected by prior expectations and beliefs. Two experiments in which prior expectations were manipulated were designed to determine whether and to what extent polygraph examiners are affected by their prior expectations when analyzing and interpreting polygraph charts. Prior expectations affected the examiners' judgments when the polygraph charts did not include clear indications of guilt or innocence, but when the objective physiological evidence included strong indications which clearly contradicted the examiner's expectations, judgments were not affected by these expectations. Theoretical and practical implications of these results are discussed.

KEY WORDS Hindsight bias Hypothesis-confirmation bias Outcome knowledge Polygraph Prior expectations Control question technique

Detecting deception and discriminating between truthful and deceptive individuals are extremely important goals in modern society. They are important both from the criminal justice perspective (e.g. to find out whether a given person is deceptive regarding a specific, usually criminal, event) and for personnel selection purposes (e.g. to discriminate between honest and dishonest individuals among a group of job applicants). It is, thus, not surprising that scientists and forensic experts have been attempting for many years to develop instruments and methods for the purpose of detecting deception.

One approach to this problem, which has received considerable attention, is psychophysiological. Various methods of psychophysiological detection of deception have been developed since the beginning of this century (e.g. Ben-Shakhar and Furedy, 1990; Lykken, 1981; Raskin, 1989; Reid and Inbau, 1977), but the one method which is clearly the most common in field practice is the Control Question Technique (CQT).

Detailed descriptions of this method can be found in various sources (e.g. Reid and Inbau, 1977; Raskin, 1989), and therefore we shall provide here only a brief description of the CQT, as typically used in the criminal investigation context. The CQT is administered in several stages: First, the

CCC 0894–3257/94/040279–14

Received 11 January 1994
Accepted 26 April 1994

examiner becomes familiar with the facts of the case by reading the written report and by speaking directly to the police investigator who ordered the examination. Typically, relevant background information, such as the suspect's past criminal record, is made available to the examiner. During the next stage the examiner conducts an extensive pre-test interview in which the examinee is given the opportunity to talk about the offense and to present his or her version of the case. The series of questions, to be asked later in the actual examination stage of the polygraph test, is formulated during this pre-test interview through an interaction between the examiner and the examinee. 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. The examiner explains the testing procedure and informs the examinee that the examination is voluntary. The next stage is the actual examination stage during which the examinee is attached to the polygraph.

During the examination stage a series of questions is presented to the examinee while continuously measuring the various physiological reactions. The questions are of the following three general types:

(1) Relevant questions — directly crime-relevant questions of the 'Did you do it?' type (e.g. 'Did you break into Mr Jones's apartment last Friday night?').
(2) Control questions — focusing on general, non-specific misconducts, of a nature as similar as possible to the issue under investigation (e.g. 'Have you ever taken something that did not belong to you?').
(3) Irrelevant questions — focusing on completely neutral issues (e.g. 'Are you sitting on a chair?').

These are intended to absorb the initial orienting response evoked by any opening question, and to enable rest periods between the more loaded questions. Typically, the whole question series is repeated three or four times. The inference rule underlying the CQT is based on a comparison of the responses evoked by the relevant and the control questions. Deceptive individuals are expected to show more pronounced responses to the relevant questions, whereas truthful individuals are expected to show the opposite pattern of responsivity (i.e. more pronounced responses to the control questions).

The CQT raised a major controversy, revolving around its rationale and inference rule, as well as around the empirical question of its validity (e.g. Ben-Shakhar and Furedy, 1990; Furedy and Heslegrave, 1989; Lykken, 1974, 1978; Raskin, 1982, 1989). In addition, some researchers have argued that the polygraph examiner knows a great deal more than what is revealed through the physiological data gathered during the examination (e.g. Ben-Shakhar *et al.*, 1986; Ben-Shakhar and Furedy, 1990). From our description of the typical CQT examination it is clear that CQT polygraph examiners are exposed to a great deal of non-physiological information, such as information provided to them by the investigator and impressions formed during the pre-test interview and during the test itself. It is impossible to differentiate between the impression formed by this prior information and those gained from the purely physiological data obtained during the test phase of the polygraph examination procedure. This feature, which has been labeled by Ben-Shakhar *et al.*, (1986) as 'contamination', implies that judgments and conclusions derived from the physiological information are contaminated with various kinds of non-physiological information. Contamination is inherent to the CQT, because this procedure is not limited to the psychophysiological data, but rather relies on the whole examiner examinee interaction, including the pre-test interview. Furthermore, CQT polygraphers believe that it is essential that the same examiner construct the questions during the pre-test interview and administer them during the test phase of the examination. Typically, the same person evaluates the polygraph charts and draws the final conclusion on the basis of all the available information, although in some polygraph agencies (including the Israeli Police) the procedure includes an additional, 'blind', evaluation of the polygraph charts. One implication of the contamination feature of the CQT is

that the weight of the strictly physiological information in the polygraph examiner's conclusion is not known, and in principle can be very small.

It should be noted that contamination does not necessarily lead to an increased error rate and a decreased validity. In principle, the non-physiological information can be accurate and valid, and polygraph examiners are trained to utilize all the information they can in an attempt to improve the accuracy of their conclusions. However, contamination may introduce various biases, because the prior information may affect the formulation of the questions and the way they are presented to the suspects. For example, when examiners believe that a given examinee is deceptive, they may present the relevant questions in a way that may affect the results in the expected direction. On the other hand, when examiners are under the impression that their suspect is truthful, the control questions might be overemphasized. In addition, the prior information may affect the judgments of polygraph examiners even when they evaluate the polygraph charts in themselves. This is a particularly likely possibility because the procedure of chart evaluation is often subjective, and precise quantification rules are not available in many polygraph agencies.

Ben-Shakhar (1991) raised the possibility that judgments of CQT examiners are affected by prior expectations. He hypothesized that polygraph examiners often generate a hunch regarding the veracity of their examinee, on the basis of the non-physiological information that was available to them (e.g. the pre-test interview, and information gathered from previous, non-polygraphic interrogations). The test-specific information is then used to test this prior hypothesis, but the hypothesis-testing process is influenced by a confirmation bias or by a primacy effect (i.e. when the judge is supposedly considering the 'objective' data neutrally for the purpose of diagnosis, or evaluation, he or she is in fact searching it for confirmation of the initial impression or the prior hypothesis which the judge entertains before looking at the data). Some data sources are sufficiently simple or well defined that they hardly lend themselves to variable interpretation. However, in the case of CQT-polygraphy, where the rich and complex physiological data are subjectively evaluated by examiners, rather than mechanically quantified, it can readily be distorted if the search process is biased and precommitted. Moreover, under these circumstances a biased search is likely to produce supportive findings — especially if it is untempered by critical attempts to falsify the initial hypothesis or to pit it against some competing alternative. The richness and vagueness of the information increase the likelihood of finding some confirmatory evidence. Indeed, it is possible that most polygraph charts contain some confirmatory information for almost any possible hypothesis.

The literature on human judgment and decision making describes several biases that might be relevant to judgments made by CQT polygraphers. For example, Evans (1989) mentions belief bias (i.e. the effect of prior beliefs and attitudes on reasoning and judgment), hindsight bias (i.e. the effect of outcome knowledge on estimates of *a priori* probabilities of events), and confirmation bias (i.e. the tendency to seek information consistent with one's current belief or hypothesis). The hypothesis formulated by Ben-Shakhar (1991) is based on a combination of the belief bias and the confirmation bias, because it postulates that polygraph examiners typically develop a belief or hypothesis based on information gathered before the administration of the CQT, and that the interpretation of the charts is influenced by that prior belief through a focus on physiological information which is consistent with it.

As noted by Klayman and Ha (1987), the term 'confirmation bias' appears in the literature with different meanings. Our use of this term differs from its original use by Wason (e.g. 1960, 1968) and his followers. In the context of polygraph chart interpretation, all the physiological information is available to the examiner, and the optimal use of this information is by proper weighting of all the cues, whether confirming the initial hypothesis or not. The confirmation bias was demonstrated mainly in the context of social perception and interaction (e.g. Darley and Gross, 1983; Snyder

and Swann, 1978a,b). These demonstrations utilized artificial experimental procedures with university students serving as subjects, and therefore it is not clear whether their results would generalize to situations where experts (e.g. polygraph examiners) are performing familiar tasks in realistic situations.

There are some demonstrations showing that judgments of experts may be affected by certain types of biases. For example, Detmer *et al.* (1978), showed that judgments of surgeons are affected by the hindsight (i.e. outcome knowledge), and Arkes *et al.* (1981) reported a similar bias with physicians. Stenson and Kleinmuntz (1975) demonstrated that experts were biased by prior instructions when sorting MMPI profiles. On the other hand, Smith and Kida (1991) suggested that biases in human judgment may have been overgeneralized, and people can be effective judges when operating in natural, familiar contexts. Christensen-Szalanski and Bushyhead (1981) produced mixed results and demonstrated that physicians overestimated the patients' probability of pneumonia but were sensitive to relative differences in the predictive value of symptoms, and appeared to use base-rate information correctly when making clinical judgments.

The main purpose of the present study is to determine whether and to what extent prior expectations affect judgments and decisions made by professional CQT-polygraph examiners working in natural, realistic setups. Two experiments in which prior expectations were manipulated by providing some examiners with outcome knowledge (e.g. telling them that the suspect had eventually confessed) were conducted. In these experiments, which focused on the effect of outcome knowledge on chart evaluation and interpretation, examiners were asked to score and assess polygraph charts from previous examinations. These two experiments are similar in many ways to studies that dealt with the hindsight bias (e.g. Fischhoff, 1975; Slovic and Fischhoff, 1977), and demonstrated that reporting the outcome of a historical event or a scientific experiment increased the perceived likelihood of that outcome.

EXPERIMENT 1

Method

Subjects

Subjects were ten polygraph examiners employed by the Israel National Police. Six of them had at least 3 years of experience in scoring polygraph records and the other four had about one.

Stimulus material

The first three charts (i.e. three repetitions of the question list) of 14 real-life criminal polygraph records were selected to be used in the present study. All records resulted from CQT polygraph examinations conducted by field-trained examiners who used Lafayette field model polygraphs, which recorded the following physiological measures:

(1) Dual respiration (thoracic and abdominal), recorded by pneumatic tubes positioned around the thoracic area and abdomen.
(2) Skin-resistance responses (SRR), recorded by two stainless steel electrodes attached to the volar side of the index and fourth fingers of the subject's left hand.
(3) Cardiovascular activity, recorded by a pneumatic blood pressure cuff positioned around the upper right arm, and inflated to a pressure between 40–50 mm Hg.

All polygraph records chosen for this study indicated inconclusive results.[1] To determine the inconclusive nature of the records, they were scored blindly by three independent examiners, who were not scheduled to take part in the experimental phase. Each of these examiners had 15 years' experience in scoring polygraph records, and therefore their scoring may serve for comparative purposes as a control for the experimental conditions. The global scores assigned by these control examiners were averaged, and it turned out that all 14 averages were within the 'inconclusive zone' (i.e. between -5 and $+5$), although in two cases one of the three scorers assigned a global score larger than $+5$. To estimate the interscorer reliability, a Pearson correlation coefficient was computed between the global scores assigned by each pair of independent scorers. The three coefficients were 0.66, 0.66, and 0.78. Typically, reliabilities among polygraph examiners using numerical scoring procedures are much higher (e.g. Raskin, 1989). However, in this experiment interrater correlations were computed on the basis of a restricted range because only inconclusive charts were used, and this might account for the attenuation in the observed correlations.

Procedure

The 14 records were arbitrarily divided into two sets of seven records each. Each set was distributed to the ten polygraph examiners for a blind numerical scoring. The examiners were requested to score the charts at their own pace. To manipulate the examiners' expectations, each group of records was accompanied by different outcome information: In the Guilt-expectation condition the examiners were told that the examinee ultimately confessed to being responsible for the crime, while in the Innocence-expectation condition they were informed that another person had confessed to that crime. Five polygraph examiners scored one set of seven records under the Guilt-expectation condition, while the other five examiners scored the same set of records under the Innocence-expectation condition. The order of the two conditions was counterbalanced across examiners.

To increase the credibility of the instructions, all the examiners participating in this experiment were told that some of the records they were asked to score were mistakenly interpreted by the original examiner. An analysis of these mistakes revealed that some of them might have been prevented had the original examiner considered only the first three charts of each examinee and avoided the fourth chart. The examiners were further told that the purpose of the study was to determine whether reducing the records to three charts would increase the accuracy of blind post-test interpretations.

The scoring was blind with respect to the nature of the case, to the background information about the case, to the specific questions presented in the polygraph test, and to the test's results. The only information provided to the examiners was about the type of each question (i.e. relevant, control, or irrelevant). However, as part of the experimental manipulation, examiners were provided with information about the guilt or innocence of the examinee (i.e. whether the examinee had confessed,

[1] The process of evaluating polygraph charts at the Polygraph Unit of the Israeli Police Force follows the numerical scoring procedure which was proposed originally by Backster (1963). According to this procedure, two or three pairs of Relevant-Control questions are identified in each polygraph chart, and numbers (-3, -2, -1, 0, 1, 2, 3) are assigned to each pair for each physiological measure. The absolute value of the assigned number reflects the magnitude of the difference between the responses evoked by the two questions within the pair (e.g. -3 or $+3$ reflect a very large difference, -1 or $+1$ reflect a small difference and 0 reflects no difference), and the sign of the assigned number reflects the direction of the difference, such that positive numbers are associated with a pattern of larger physiological reactivity to the control question, and negative numbers reflect the opposite pattern. These numbers are then summed up across question pairs, across physiological measures, and across polygraph charts to yield a global score. Thus, if for example a polygraph examination is based on three charts and three physiological measures and if two pairs of Relevant-Control questions are identified for each chart, then the global score ranges between -54 and $+54$. Typically the following classification rule is used. If the global score exceeds $+5$, an NDI (no deception indicated) classification is reached: if the global score is less than -5, the polygraph record is classified as DI (deception indicated); and if the global score ranges between -5 and $+5$, the record is classified as inconclusive.

Exhibit 1. Frequencies of the classifications made by each examiner under the two experimental conditions and the control condition of Experiment 1

Experimental condition	Guilt expectation			Innocence expectation			Control		
Examiner	NDI	Inclusive	DI	NDI	Inclusive	DI	NDI	Inclusive	DI
1[a]	0	6	1	4	3	0			
2[a]	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[a]	1	6	0	0	7	0			
7[a]	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			
11							1	13	0
12							1	13	0
13							0	14	0
Across examiners	6	61	3	15	55	0	2	40	0

NDI = no deception indicated; DI = deception indicated.
[a] Less experienced examiners.

or whether he or she was exonerated through the confession of another person suspected of committing the same crime).

Results

Each polygraph record was analyzed by each of the ten examiners using the numerical scoring procedure. The global scores assigned to the polygraph records were classified into three categories: DI (deception indicated), if the global score was less than or equal to −6; NDI (no deception indicated), if the global score was greater than or equal to +6; and inconclusive if the global score was between −5 and +5. Exhibit 1 presents the frequencies of the three classifications made by each examiner under each experimental condition, as well as the classifications made by the three control examiners under the control condition of no outcome knowledge.

Exhibit 1 reveals that under the Innocence-expectation condition 21% (15 out of 70) of the polygraph records were classified as NDI, whereas none was classified as DI. In comparison, the frequency of NDI classifications under the control condition was only 5% (two out of 42). Under the Guilt-expectation condition, 8.6% and 4.3% of the records were classified as NDI and DI, respectively. The distribution of these classifications does not allow for statistical tests (the vast majority of the cases were classified as inconclusive, and the frequencies of the other categories are much too small). Therefore we conducted statistical tests on the continuous scale of the global scores. Because different examiners might differ in their scoring tendencies, a within-examiner standardization of the global scores was used (Ben-Shakhar, 1985). Each global score was transformed into a standard score relative to the mean and the standard deviation of the individual examiner's global score distribution.

The average Z scores across examiners within each experimental condition was computed for each polygraph record. Exhibit 2 displays the mean Z scores for each polygraph record under each of the two experimental conditions and under the control condition. For 13 out of the 14 records, the difference between the average Z scores was in the expected direction (i.e. more positive scores under the Innocence-expectation condition than under the Guilt-expectation condition). A matched-

Exhibit 2. Mean standard scores computed for each polygraph record across examiners under the two experimental conditions and the control condition of Experiment 1

Experimental condition	Guilt expectation	Innocence expectations	Control
Record			
1	−0.51	−0.00	+0.25
2	−1.04	−0.62	−0.80
3	−0.93	−0.15	−0.73
4	+0.66	+1.25	+0.85
5	−0.17	+1.19	+1.39
6	−0.86	+0.24	−0.83
7	−0.37	+0.02	−1.30
8	+0.14	+0.91	+0.52
9	−0.43	−0.06	+0.36
10	+0.73	+1.08	+1.52
11	−1.37	+0.72	−0.95
12	−0.46	+0.52	−0.74
13	+0.49	+0.28	+0.99
14	−1.03	−0.22	+0.54
Marginal mean	−0.368	+0.369	+0.076

group *t*-test was conducted to compare the mean *Z* scores obtained under the Innocence-expectation condition with those obtained under the Guilt-expectation condition. This comparison yielded a statistically significant difference[2] ($t_{(13)} = 5.04$).

A similar analysis was conducted for the examiners. A mean *Z* score was computed for each of the ten examiners, under each experimental condition across the seven polygraph records scored by the examiner. These means, which were computed separately for the experienced and for the less experienced examiners, are presented in Exhibit 3. The *Z* scores obtained under the two experimental conditions were compared separately for the experienced and inexperienced examiners using matched-group *t*-tests. The differences were statistically significant for both the experienced examiners ($t_{(5)} = 3.05$), and for the inexperienced examiners ($t_{(3)} = 3.90$).

A closer inspection of Exhibit 3 reveals that Examiner 7, a less experienced examiner, was highly influenced by outcome knowledge (there was no overlap between the scores assigned by this examiner under the two experimental conditions), whereas Examiner 4, a more experienced examiner, was not affected by the expectations at all. It is possible that examiners differ in their vulnerability to this 'hindsight bias' (e.g. Fischhoff, 1982). However, excluding these two extreme cases, variations among examiners seem to be negligible, and the data suggest that experience cannot account for the expectation effect observed in this experiment.

Discussion

The results of this experiment revealed that outcome knowledge affected the examiners in their scoring of CQT records. Although the effect was reliably demonstrated only with the continuous numerical scale, and not with the classification into broad categories, it seems to be a rather strong effect, as it reached accepted levels of significance with a relatively small sample size. This means that polygraph examiners are affected by prior beliefs when they analyze and interpret polygraph charts. Moreover, prior expectations affected both experienced and inexperienced examiners. This is consistent with Wood (1978), who found that the most knowledgeable subjects were no less prone to hindsight

[2] A significance level of 0.05 was used in all statistical analyses reported in this study.

Exhibit 3. Mean standard scores computed for experienced and less experienced polygraph examiners under the two experimental conditions of Experiment 1

Experimental condition	Guilt expectation	Innocence expectations
Experienced examiners		
1	−0.44	+0.44
2	−0.44	+0.43
3	−0.12	+0.13
4	+0.09	−0.09
5	−0.44	+0.44
6	−0.27	+0.27
Less experienced examiners		
7	−0.92	+0.92
8	−0.39	+0.39
9	−0.44	+0.44
10	−0.32	+0.32
Marginal mean	−0.369	+0.369

bias than less knowledgeable ones in a task involving general-knowledge questions. Fischhoff (1982) concluded that people normally approach hindsight tasks in a manner that does not use their knowledge or inferential skills at an optimal level.

In the first experiment, only inconclusive records were used, and this raises a question regarding the generalizability of the present results to situations where the actual physiological responses clearly indicate deception or innocence. In other words, it is possible that prior information and prior expectations affect polygraph examiners' judgment only when the specific information provided by the polygraph charts is unclear or ambiguous. It is thus important to determine whether outcome knowledge affects examiners' judgment when it stands in complete contrast to the physiological information. It was therefore decided to replicate the first experiment using conclusive polygraph records.

EXPERIMENT 2

The second experiment was similar to the first in every respect, with one exception: it utilized conclusive rather than inconclusive polygraph records.

Method

The same ten examiners were asked for a second opinion about several records used in the first experiment for reliability purposes. The examiners were further told that several new records would also be presented. In fact, all records were new records. Six of them were selected because they clearly indicated the innocence of the subjects (a mean global score of +6 or more was assigned by two independent and experienced scorers who were not scheduled to participate in the experiment). The other six records clearly indicated guilt (a mean global score of −6 or less was assigned to them by these two experienced examiners). The interscorer correlation coefficient computed for the global score was 0.88.

After completion of this task and another task (not reported here), the polygraph examiners were debriefed and the real purpose of the study was disclosed. Their reaction was of surprise, but no one expressed anger, disapproval, or other negative feelings. The examiners acknowledged the import-

Exhibit 4. Frequencies of classifications made by ten polygraph examiners under two experimental conditions and the control condition of Experiment 2

Experimental condition	Innocence expectation		Guilt expectation		Control	
Actual Outcome	NDI	DI	NDI	DI	NDI	DI
Classification						
NDI	28	0	29	0	12	0
Inconclusive	2	9	1	7	0	2
DI	0	21	0	23	0	10
	30	30	30	30	12	12

NDI = no deception indicated; DI = deception indicated.

ance of the study for achieving a better understanding of their profession, and all of them understood that it would have been impossible to conduct the study without deception.[3]

Results

As in Experiment 1, all polygraph records were classified into the three categories of DI, NDI, and Inconclusive, based on the global scores assigned to them by the ten examiners. Exhibit 4 presents the frequencies of the three categories made under the two experimental conditions, as well as under the control condition of no outcome knowledge.

Exhibit 4 indicates that the classifications based on the global scores assigned by the ten examiners reflect the actual outcomes of the records and are unrelated to the experimental condition.

The global scores were standardized within examiners as in Experiment 1, and the average Z score across examiners within each experimental condition was computed for each polygraph record. Exhibit 5 displays the mean Z scores for each polygraph record under each experimental condition. For seven out of the 12 records the difference between the average Z scores was in the expected direction (i.e. larger mean Z scores under the Innocence-expectation condition than under the Guilt-expectation condition). The mean difference across all records was in the expected direction, but a matched-group t-test revealed that it was not statistically significant ($t_{(11)} = 1.42$).

A similar analysis was conducted for the examiners. A mean Z score was computed for each of the ten examiners, under each experimental condition, across the six polygraph records scored by the examiner. These means are presented in Exhibit 6. The differences between the Z scores obtained under the two experimental conditions were in the expected direction for eight of the ten examiners. However, these differences were small and a matched-group t-test indicated that they were not statistically significant ($t_{(9)} = 1.25$).

Discussion

Unlike the previous experiment, the results of Experiment 2 failed to produce a statistically significant effect for the examiners' expectations, although the differences were in the predicted direction. It seems that when prior expectations clearly contradict the specific, physiological information provided by the polygraph charts, examiners tend to rely primarily upon the latter and pay little attention

[3] There are no formal ethical guidelines for conducting research in the Israel Police. However, as research psychologists, we are committed to the ethical guidelines of APA, according to which deception should be avoided unless it is scientifically essential to the study, and it is estimated that the subjects will not be distressed by the deception when debriefed. In this case deception was absolutely necessary and the examiners expressed surprise but not anger or distress when debriefed.

Exhibit 5. Mean standard scores computed for each polygraph record across examiners under the two experimental conditions and the control condition of Experiment 2

Experimental condition	Guilt expectation	Innocence expectation	Control
Record			
1	+1.16	+1.30	+1.09
2	+0.62	+0.60	+0.68
3	+1.16	+1.03	+1.09
4	−0.83	−0.35	−0.66
5	−1.13	−1.05	−1.12
6	−1.12	−0.90	−1.12
7	+0.83	+1.31	+1.25
8	+0.73	+0.54	+0.68
9	+1.17	+0.83	+0.99
10	−0.98	−0.34	−1.12
11	−0.99	−1.19	−0.65
12	−1.38	−1.01	−1.01
Marginal mean	−0.063	+0.064	+0.008

Exhibit 6. Mean standard scores computed for experienced and less experienced polygraph examiners under the two experimental conditions of Experiment 2

Experimental condition	Guilt expectation	Innocence expectations
Experienced examiners		
1	−0.15	+0.15
2	−0.02	+0.02
3	+0.13	−0.13
4	−0.07	+0.07
5	−0.05	+0.05
6	−0.08	+0.08
Less experienced examiners		
7	−0.14	+0.14
8	−0.08	+0.08
9	−0.02	+0.02
10	+0.11	−0.11
Marginal mean	−0.037	+0.037

to their expectations. Thus, the results of Experiment 2 suggest that the effect of prior expectations on polygraph charts' reading is limited to inconclusive records. It may be argued along the hypothesis suggested by Ben-Shakhar (1991) that when analyzing polygraph charts, the examiner does search for confirmatory information, but when no confirmatory signs exist, the hypothesis generated from the prior expectations is rejected, and the final judgment reflects the physiological information. It should be noted that polygraph examiners at the Israel Police are trained to score polygraph records using the semi-objective technique suggested by Backster (1963). It is possible that examiners trained by more traditional schools, which emphasize subjective methods of global evaluation of polygraph records, would be more vulnerable to biases of prior beliefs and expectations, and would be affected by expectations even when scoring polygraph records that include contradictory signs.

It cannot be completely ruled out that the experimental manipulation was not sufficiently strong,

in which case the relatively large discrepancies between the charts and the prior expectations might have raised some suspicion regarding the credibility of the alleged confessions.

GENERAL DISCUSSION

The present study was conducted to determine the impact of prior expectations on the judgments of polygraph examiners when analyzing polygraph charts obtained from a CQT polygraph examination. The influence of prior expectations on human judgment has been demonstrated mainly in artificial experimental conditions (e.g. Fischhoff, 1982; Snyder and Swann, 1978a,b), but there are some demonstrations that it can affect professionals working in more realistic and familiar situations (e.g. Arkes *et al.*, 1981; Detmer *et al.*, 1978). Ben-Shakhar (1991) argued that judgments of CQT-polygraph examiners may be particularly vulnerable to a bias created by prior expectations and beliefs, because the CQT is based to a large extent on the examiner–examinee interactions, rather than on objective inference rules and quantification schemes. He further argued that the interpretation of the complex physiological information is infected by a confirmation bias (i.e. larger weight is given to physiological data consistent with the prior expectations, while contradictory evidence is largely ignored). Furthermore, it was argued that prior expectations can affect not only the evaluation of polygraph charts but also the choice of the control questions and the manner in which they are administered to the examinees.

The results of the present experiments supply only partial support to the hypotheses raised by Ben-Shakhar (1991). First, prior expectations had an impact on the examiners' judgments when they analyzed previously obtained polygraph charts. The impact of prior expectations on polygraph examiners conducting the entire CQT test and having access not only to the physiological responses but also to the subjects' behavior during the various phases of the CQT test will have to be examined in future studies. To use Darley and Gross (1983) terminology, only the 'cognition confirmation effect' (i.e. expectancy-confirmation effects that occur in the absence of any interaction between the perceiver and the target person) was demonstrated in this study, but not the 'behavioral confirmation effect' (i.e. where expectations affect the behavior toward a target individual, such that expectancy-confirming behaviors are elicited from this individual).

Second, the impact of prior expectations on the interpretation of polygraph charts was observed only when these charts did not include clear indications of either guilt or innocence. In other words, it seems that when the specific physiological information clearly contradicts prior expectations, examiners tend to ignore their expectations and make their judgments on the basis of the physiological information. This finding is consistent with the interpretation of hindsight effects in terms of availability (i.e. failure to consider alternative outcomes). Slovic and Fischhoff (1977) suggested that victims of hindsight bias focus their attention on the reported outcome and fail to consider alternatives. Indeed, they demonstrated that once subjects were encouraged to search for possible explanations for the alternative outcome (the outcome that did not happen), the hindsight effect was reduced. A similar account was provided by Koriat *et al.* (1980) to explain overconfidence in human judgment. They suggested that overconfidence occurs because subjects are attempting to justify their chosen answer, and in this process they focus on evidence which supports this answer, and disregard contradictory evidence. Koriat *et al.* (1980) demonstrated that inducing subjects to list contradicting reasons reduced the overconfidence effect. The results of Experiment 2 suggest that providing examiners with physiological information which clearly contradicts their prior expectations drastically reduced the bias. These clear records are bound to bring the alternative possibility to the attention of the polygraph examiners, and thus reduce the effect of outcome knowledge, and possibly eliminate it.

The demonstration of a prior-expectations effect on professional polygraph examiners may have

some practical implications. Even if such an effect operates only for some polygraph examiners and only when the physiological information is not conclusive, it might increase the error rate of classifications made by polygraph examiners. Several measures can be suggested in order to minimize and possibly eliminate the effects of prior expectations in CQT polygraph examinations. First, a procedure of blind scoring of the charts (i.e. scoring by an examiner who is unaware of the details of the case under investigation, and is unfamiliar with the suspect) should be routinely implemented. Such a procedure has been adopted already by the Israel Police, and when there are disagreements between the original examiner and the blind scorer, the case is further discussed. Ideally, charts should be scored mechanically by a computer, using various techniques for measuring physiological reactions that are available in experimental psychophysiology (e.g. Kircher and Raskin, 1988). In principle, even a complete computerization of physiological measurement may be insufficient to eliminate all possible sources of bias, because prior beliefs and expectations may affect not only chart interpretation but also the whole manner in which the CQT is administered. Ben-Shakhar *et al.* (1986) listed several steps that should be taken to achieve a complete decontamination of polygraph examinations, although they doubted whether such changes would be acceptable to CQT examiners who emphasize the examiner–examinee interaction. Additional measures that can reduce the impact of prior expectations are the use of a larger range of inconclusive classifications, and a greater emphasis on the dangers of relying upon prior information in the training of polygrapy examiners.

Finally, it should be noted that effects of prior expectations on human judgment may extend to other situations where experts are required to make judgments and predictions on the basis of various tools. Ben-Shakhar (1991) argued that whenever expert judgments are based on subjective tools with no well-defined quantification schemes and inference rules, and when the expert is exposed to a great deal of prior information, the judgments allegedly made on the basis of a scientific instrument might be seriously distorted by other sources of information and might reflect to a large extent the prior beliefs of the judge. For example, many of the tools commonly used for personnel selection and diagnosis (e.g. projective techniques, personal interviews) are based on subjective judgment, rather than on objective and quantified inference rules. Ben-Shakhar (1991) suggested that such tools provide the users with an impression of validity (personal validity) created through the process of confirming initial hypotheses (a test which so often confirms the user's prior expectations acquires a sense of validity). Another example is the preference for intuitive (clinical) judgment over the use of statistical models for prediction. The advantage of statistical over clinical prediction has been documented repeatedly (e.g. Dawes, 1979), yet personnel decisions are typically made at least partly on the basis of intuitive judgment. Future research should focus on these issues and determine which type of tools and testing situations are vulnerable to both the cognitive and the behavioral confirmation effects.

ACKNOWLEDGEMENTS

We wish to thank the polygraph examiners of the Israel Police for their assistance in the execution of this study. In addition, we are grateful to Maya Bar-Hillel, Gideon Keren, Benjamin Kleinmuntz, and an anonymous reviewer for their helpful comments on an earlier version of this manuscript.

REFERENCES

Arkes, H. R., Wortman, R. C., Saville, P. D. and Harkness, A. R. 'Hindsight bias among physicians weighing the likelihood of diagnosis', *Journal of Applied Psychology*, **66** (1981), 252–254.

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

Ben-Shakhar, G. 'Standardization within individuals: A simple method to neutralize individual differences in psychophysiological responsivity', *Psychophysiology*, **22** (1985), 292–299.
Ben-Shakhar, G. 'Clinical judgment and decision making in CQT polygraphy: A comparison with other pseudo-scientific applications in psychology', *Integrative Physiological and Behavioral Science*, **26** (1991), 232–240.
Ben-Shakhar, G., Bar-Hillel, M. and Lieblich, I. 'Trial by polygraph: Scientific and juridical issues in lie detection', *Behavioral Science and the Law*, **4** (1986), 459–479.
Ben-Shakhar, G. and Furedy, J. J. *Theories and Applications in the Detection of Deception: A psychophysiological and international perspective*, New York: Springer-Verlag, 1990.
Christensen-Szalanski, J. J. J. and Bushyhead, J. B. 'Physicians' use of probabilistic information in a real clinical setting', *Journal of Experimental Psychology: Human Perception and Performance*, **7** (1981), 928–935.
Darley, J. M. and Gross, P. H. 'A hypothesis-confirmation bias in labeling effects', *Journal of Personality and Social Psychology*, **44** (1983), 20–33.
Dawes, R. M. 'The robust beauty of improper linear models in decision making', *American Psychologist*, **34** (1979), 571–582.
Detmer, D. E., Fryback, D. G. and Gassner, K. 'Heuristics and biases in medical decision-making', *Journal of Medical Education*, **53** (1978), 682–683.
Evans, J. St B. T. *Bias in Human Reasoning: Causes and consequences*, London: Lawrence Erlbaum, 1989.
Fischhoff, B. 'Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty', *Journal of Experimental Psychology: Human Perception and Performance*, **1** (1975), 288–299.
Fischhoff, B. 'Debiasing', in Kahneman, D., Slovic, P. and Tversky, A. (eds), *Judgment under Uncertainty: Heuristics and biases*, pp. 422–444, Cambridge: Cambridge University Press, 1982.
Furedy, J. J. and Heslegrave, R. J. 'The forensic use of the polygraph: A psychophysiological analysis of current trends and future prospects', in Jennings, J. R., Ackles, P. K. and Coles, M. G. H. (eds), *Advances in Psychophysiology*, 4, Greenwich, CT: JAI Press, 1989.
Kircher, J. C. and Raskin, D. C. 'Human versus computerized evaluations of polygraph data in laboratory setting', *Journal of Applied Psychology*, **73** (1988), 291–302.
Klayman, J. and Ha, Y.-W. 'Confirmation, disconfirmation, and information in hypothesis testing', *Psychological Review*, **94** (1987), 211–228.
Koriat, A., Lichtenstein, S. and Fischhoff, B. 'Reasons for confidence', *Journal of Experimental Psychology: Human Learning and Memory*, **6** (1980), 107–118.
Lykken, D. T. 'Psychology and the lie detection industry', *American Psychologist*, **29** (1974), 725–739.
Lykken, D. T. 'Uses and abuses of the polygraph', in Pick, H. L. (ed.), *Psychology: From Research to Practice*, New York: Plenum Press, 1978.
Lykken, D. T. *A Tremor in the Blood: Uses and Abuses of the Lie Detector*, New York: McGraw-Hill, 1981.
Raskin, D. C. 'The scientific basis of polygraph techniques and their uses in the Judicial process', in Trankell, A. (ed.), *Reconstructing the Past: The role of psychologists in the criminal trial*, Stockholm: Norsted & Soners, 1982.
Raskin, D. C. 'Polygraph techniques for the detection of deception', in Raskin, D. C. (ed.), *Psychological Methods in Criminal Investigation and Evidence*, New York: Springer-Verlag, 1989.
Reid, J. E. and Inbau, F. E. *Truth and Deception: The Polygraph ('Lie Detection') Technique*, Baltimore, MD: Williams and Wilkins, 1977.
Slovic, P. and Fischhoff, B. 'On the psychology of experimental surprise', *Journal of Experimental Psychology: Human Perception and Performance*, **3** (1977), 544–551.
Smith, J. F. and Kida, T. 'Heuristics and biases: Expertise and task realism in auditing', *Psychological Bulletin*, **109** (1991), 472–489.
Snyder, M. and Swann, W. B., Jr. 'Behavioral confirmation in social interaction: From social perception to social reality', *Journal of Experimental Psychology*, **14** (1978a), 148–162.
Snyder, M. and Swann, W. B., Jr. 'Hypothesis-testing processes in social interaction', *Journal of Personality and Social Psychology*, **36** (1978b), 1202–1212.
Stenson, H., Kleinmuntz, B. and Scott, B. 'Personality assessment as a signal detection task', *Journal of Consulting and Clinical Psychology*, **43** (1975), 794–799.
Wason, P. C. 'On the failure to eliminate hypotheses in a conceptual task', *Quarterly Journal of Experimental Psychology*, **12** (1960), 129–140.
Wason, P. C. 'On the failure to eliminate hypotheses — a second look', in Wason, P. C. and Johnson-Laird, P. N. (eds), *Thinking and Reasoning*, Penguin, Harmondworth, 1968.
Wood, G. 'The knew-it-all-along effect', *Journal of Experimental Psychology: Human Perception and Performance*, **4** (1978), 345–353.

Authors' biographies:
Eitan Elaad is the research officer of the Scientific Interrogation Unit, Israel Police. He is also a lecturer at the Department of Criminology, Bar-Ilan University. He received his PhD in Psychology from the Hebrew University of Jerusalem in 1988. Research interests include psychophysiological detection of concealed knowledge, the social psychology of deception, and decision-making processes.
Avital Ginton is a commander in the National Police of Israel and Head of Scientific Interrogation and Behavior Section. He received a PhD in Psychology from Tel-Aviv University in 1978, and an MPA from Harvard University in 1993. For many years he lectured in the Department of Psychology at Tel-Aviv and Criminology at Bar-Ilan Universities.
Gershon Ben-Shakhar (PhD, Hebrew University of Jerusalem) is a Professor of Psychology at the Hebrew University of Jerusalem. His main research is focused on human psychophysiology (orientation and habituation processes), and psychophysiological detection of information. He also studied psychological testing and its applications to personnel decisions.

Authors' addresses:
Eitan Elaad and **Avital Ginton**, Behavior Section, Division of Identification and Forensic Science, Israel National Police Headquarters, Jerusalem, 91906 Israel.
Gershon Ben-Shakhar, Department of Psychology, The Hebrew University of Jerusalem, Israel.