Psychophysiology

Vol. 28, No. 5
Printed in U.S.A.

The Truth Will Out: Interrogative Polygraphy ("Lie Detection") With Event-Related Brain Potentials

Lawrence A. Farwell and Emanuel Donchin
Cognitive Psychophysiology Laboratory, University of Illinois at Urbana-Champaign

ABSTRACT

The feasibility of using Event Related Brain Potentials (ERPs) in Interrogative Polygraphy ("Lie Detection") was tested by examining the effectiveness of the Guilty Knowledge Test designed by Farwell and Donchin (1986, 1988). The subject is assigned an arbitrary task requiring discrimination between experimenter-designated targets and other, irrelevant stimuli. A group of diagnostic items ("probes"), which to the unwitting are indistinguishable from the irrelevant items, are embedded among the irrelevant. For subjects who possess "guilty knowledge" these probes are distinct from the irrelevants and are likely to elicit a P300, thus revealing their possessing the special knowledge that allows them to differentiate the probes from the irrelevants. We report two experiments in which this paradigm was tested. In Experiment 1, 20 subjects participated in *one* of two mock espionage scenarios and were tested for their knowledge of *both* scenarios. All stimuli consisted of short phrases presented for 300 ms each at an interstimulus interval of 1550 ms. A set of items were designated as "targets" and appeared on 17% of the trials. Probes related to the scenarios also appeared on 17% of the trials. The rest of the items were irrelevants. Subjects responded by pressing one switch following targets, and the other following irrelevants (and, of course, probes). ERPs were recorded from F_Z, C_Z, and P_Z. As predicted, targets elicited large P300s in all subjects. Probes associated with a given scenario elicited a P300 in subjects who participated in that scenario. A bootstrapping method was used to assess the quality of the decision for each subject. The algorithm declared the decision indeterminate in 12.5% of the cases. In all other cases a decision was made. There were no false positives and no false negatives: whenever a determination was made it was accurate. The second experiment was virtually identical to the first, with identical results, except that this time 4 subjects were tested, each of which had a minor brush with the law. Subjects were tested to determine whether they possessed information on their own "crimes." The results were as expected; the Guilty Knowledge Test determined correctly which subject possessed which information. The implications of these data both for the practice of Interrogative Polygraphy and the interpretation of the P300 are discussed.

DESCRIPTORS: Event-related potentials, Lie detection, Guilty Knowledge Test, P300.

We report an examination of the feasibility of adopting an approach to Interrogative Polygraphy[1] (Farwell & Donchin, 1986, 1988, 1989) that bases inferences regarding the degree to which subjects possess knowledge they are hiding on mea-

The research reported here was supported in part by contract number 87F350800 with the Central Intelligence Agency. Preliminary reports were presented at the 1986, 1988, and 1989 meetings of the Society for Psychophysiological Research. We acknowledge with thanks the comments of Bill Gehring, Greg Miller, and David Lykken on earlier versions of this manuscript. The project could not have been conducted without the technical support of the staff of the Cognitive Psychophysiology Laboratory (CPL). Special thanks are due to Brian Foote for his programming, and to Mike Anderson for his technical wizardry. Many of our CPL colleagues have helped with the design of these experiments and their interpretation as they unfolded. Particular thanks are due to Christine Collins and Gregory Miller.

Address request for reprints to: Emanuel Donchin, Department of Psychology, University of Illinois, 603 East Daniel, Champaign, Illinois 61820.

[1]We use the term Interrogative Polygraphy to refer to a body of techniques that is popularly known as "lie detection," and commonly referred to by its practitioners as "polygraphy." The popular term is inaccurate and misleading, because none of the techniques actually detect lies (see Ekman, 1985). The term polygraphy is too broad if left unqualified. All psychophysiologists use polygraphs. This paper is concerned with those psychophysiologists who are using polygraphs as aids in interrogations, and that is why the term interrogative polygraphy is deemed appropriate.

sures of attributes of various components of the event-related brain potential (ERP). This approach is based on the premise that some ERP components are manifestations, at the scalp, of the activity of subsystems that execute specific information-processing tasks (Donchin, 1981; Donchin & Coles, 1988; Donchin, Ritter, & McCallum, 1978; Hillyard & Kutas, 1983).

Although the use of ANS (autonomic nervous system) responses in Interrogative Polygraphy has many advocates (Barland & Raskin, 1975; Office of Technology Assessment, 1983; Podlesny & Raskin, 1978; Raskin, 1986; Reid & Inbau, 1977) and the technique is widely employed by government agencies and by private practitioners, it suffers a number of inherent difficulties (Ben-Shakhar & Furedy, 1990; Furedy, 1986; Lykken, 1981). It is worthwhile to examine the feasibility of augmenting the tool kit available to Interrogative Polygraphy. It should be clear, however, that the study reported here represents but the first stage in the development of an interrogatory technique. Our purpose is to demonstrate that it is possible, *at least under laboratory conditions*, to use ERPs in the detection of concealed information. This demonstration, if persuasive, should lead to further research that will examine the efficacy of the procedure under field conditions.

The Interrogative Polygraphy procedure we examine in this report utilizes the P300 component of the ERP in the context of the "oddball" paradigm (Fabiani, Gratton, Karis, & Donchin, 1987). The subject is presented with a series of events and with a classification rule that places each event into one of two categories. The classification rule can range from the concrete to the abstract (i.e., frequency differences of tones, exemplars of two different categories). The events can be quite diverse, and the categorization may depend on an ensemble of properties of the events. The subject must be assigned a task that requires the categorization of the events. The series is so constructed that events from the two categories occur in a random sequence. Furthermore, one of the categories occurs less frequently than the other. When these conditions are satisfied, the events that belong in the rare category elicit a large P300. The amplitude of the P300 is inversely proportional to the probability of the eliciting event-category and directly proportional to the relevance of the event to the subject's task (Donchin, Karis, Bashore, Coles, & Gratton, 1986; Duncan-Johnson & Donchin, 1977; Squires, Squires, & Hillyard, 1975).

Thus, the appearance of a P300 in response to the events that belong to the rare category in an oddball series indicates that the subject has correctly categorized them. This feature of the oddball paradigm suggests its use in Interrogative Polygraphy. One constructs a series whose elements would appear homogeneous to the innocent, whereas for the guilty (who are concealing information) they fall into two categories. If the critical items, which are distinct from the rest of the series by virtue of their association with the crime, elicit a P300, then one may conclude that the subject does make the discriminations that only the guilty can make. It is critical to note that for the P300 to be elicited the subject must be assigned some task whose performance depends on the processing of the events. If a two-category series appears homogeneous to the innocent, no task can be assigned unless one directs the subject's attention to the very items that are in principle indiscriminable, thus defeating the purpose of the test. Farwell and Donchin (1986, 1988, 1989) solved this problem by creating an oddball series consisting of three subseries. One subseries included stimuli that occurred 17% of the time and were defined as "targets" by virtue of their inclusion on a list of items that the subject was instructed to detect. The remaining stimuli, presented on 83% of the trials, were not included in the list and were, therefore, defined for the subject as "irrelevant." However, some of these irrelevant items (17% of the stimuli) were in fact "crime-relevant" or "probe" stimuli. Their relevance, though, was known only to the investigator and to the "criminal." Thus, the innocent subject is presented with a standard two-category oddball series in which the targets would be expected to elicit a P300 whereas the irrelevant stimuli would not. For the guilty, who possess the extra knowledge, the crime-relevant probes stand out as a third, rare category.

This procedure is closely related to the Guilty Knowledge Test (GKT) first proposed by Lykken (1959; see also Lykken, 1981). The GKT challenges the subject with a series of multiple choice questions, one of the choices for each question being a detail knowledge of which indicates guilt. For example, "Was the getaway car a (1) red ford, (2) yellow toyota, (3) pink honda, (4) grey chevy, or (5) white plymouth?" For the innocent none of the alternatives carries any significance. The robber presumably knows that she absconded in a yellow toyota. Electrodermal responses and, in some cases, cardiovascular and pulmonary responses are recorded, and a consistently enhanced response to the critical option is taken as an indication of guilt. The GKT provides an important control against false positives in that the subject's response to the noncritical items serves as a control, providing a baseline of the responsiveness to a question implying her guilt, even when she does not know which car was used in the robbery. Note, however, that the GKT has but a weak control for false negatives. Absence of an enhanced response to any option cannot be readily evaluated because nothing in the procedure forces the subject to respond to any of the items.

There are two noteworthy features of the procedure described in this report. By selecting an arbitrary set of stimuli to be designated as targets, which the subject must discriminate, a task is created that focuses the subject's attention in a manner that ensures the elicitation of the P300. However, this task is unrelated to the subject of the investigation. In contrast, when the oddball series is constructed explicitly to call for a discrimination between "crime-relevant" and "crime-irrelevant" items, the likelihood of false positives increases, because innocent subjects may process the explicitly relevant items differently from the way they would the irrelevant items.

The designation of arbitrary targets makes it possible to hide the crime-relevant items among the nontarget, frequent, and irrelevant events. For the innocent subjects, these items are indistinguishable from all other irrelevants, because nothing in the procedure draws the subject's attention to these items. For the "guilty" these items stand out among the frequent items, because they are associated with the crime, given the information the subject possesses. It is important to note that the decision as to the subject's possession of the guilty information is based on the appearance of differences among the frequent, nontarget items, differences that would not occur without the guilty knowledge. Thus, our procedure incorporates two different controls. The targets serve to identify the level of responsiveness to be expected from the subject to relevant items, guarding against false negatives. The irrelevant items serve as a control for the response to items that have no bearing on the subject of the interrogation. The indistinguishability of the probes and irrelevants guards against false positives.

This approach to Interrogative Polygraphy will be illustrated by two studies. In the first experiment subjects participated in a mock espionage scenario, and the P300 was used to detect guilty knowledge regarding the scenario. The subjects were also tested in a scenario of which they were "innocent." That a subject participated in a scenario was inferred from the fact that he was knowledgeable, as revealed by the P300, regarding the significant details of the scenario. Similarly, "innocence" was demonstrated by the subjects' ignorance of relevant details, as revealed by the absence of a P300 in response to the critical events. The second experiment shows that the technique can be used to detect information about "real world" events. The technique was used to detect guilty knowledge regarding minor crimes actually committed by each of 4 subjects, and lack of guilty knowledge regarding acts they did not commit.

EXPERIMENT 1

Method

Subjects

Twenty subjects (12 female), aged 19–27 years, participated in the study. All subjects were undergraduate or graduate students who were paid for their participation.

Procedure

Subjects were trained by an interactive computer program to perform one of two different mock espionage scenarios, which they then proceeded to carry out. There were two different scenarios. Each subject learned one scenario and was unaware of the existence of the other scenario. Thus, each subject could be tested on the knowledge of the scenario that he or she experienced (this was that subject's "Guilty" scenario) as well as on the scenario of which he was innocent. Each of the scenarios required the subjects to go to a specific location and meet a person with whom a password would be exchanged. The subject then asked that person for a file that had a particular designation and pertained to a specific operation. With each of the two scenarios we associated six critical details, knowledge of which could be used as indication that the subject participated in the scenario. Appendix A presents the information associated with each of the scenarios, as well as all the other stimuli used in the study.

The interactive training program consisted of a series of instructions about the critical items that the subjects were instructed to memorize and actions they were to follow. The instructions were repeated several times, and subjects were repeatedly tested on the instructions until they had responded correctly at least five times to questions regarding each of six key items. Following the training session, when the computer had established that the subject had learned the scenario to criterion, he or she was instructed to undertake the mission. In each case, the appropriate file folders were handed to the subject and he or she proceeded to the location where the information and files were exchanged with the contact.

One day after executing the scenario each subject underwent an ERP (event-related brain potential) aided interrogation designed to test whether the subject possessed "guilty" knowledge. Subjects were tested for knowledge of each of the two scenarios, the one in which they had actually participated and the other scenario of which they knew nothing. Subjects were tested in three blocks for each of the two scenarios. Blocks of Scenario 1 alternated with blocks of Scenario 2. Note that Scenario 1, of which half the subjects were guilty, was tested first for all subjects. Thus, the order of testing was counterbalanced across subjects. Two guilty knowledge tests were administered, one based on ERPs and one on skin conductance as measured by a conventional polygraph used in detection of deception.[2]

The Testing Environment

Stimuli were presented visually on a CRT under computer control, and the ERPs elicited by each stimulus were recorded and stored on tape for off-line analysis. Each stimulus consisted of two one-syllable words. Stimulus duration was 300 ms. The interstimulus interval was 1550 ms.

The subjects were told that the stimuli would be two-word phrases. Some of these phrases were arbitrarily designated as targets, and the subjects' task was to press one of two microswitches whenever they saw a target and to press another microswitch when they saw an irrelevant item. Although the instructions referred to two categories of stimuli, there were in fact three categories, because among the nontarget stimuli we included the "probes," which were phrases referring to the six critical items associated with each scenario.

For each of the two scenarios, subjects were tested in three blocks of 144 trials per block. On each trial we presented the subject with a two-word phrase on the screen. The phrases could be targets, irrelevants, or probes. The three categories were presented in a random order. The set of targets contained 6 phrases to which the subject was instructed to respond by pressing one of the switches. Each of the 6 target items was repeated 4 times in each sequence, so that the total number of target trials was 24, or 17% of the trials. The remaining 120 trials consisted of irrelevants which could be derived from one of two stimulus sets. The "true irrelevants" included the items that bore no relationship to either of the scenarios. For each target there were 4 similar irrelevants, for a total of 24 unique irrelevants. The phrases used as targets and as irrelevants are presented in Appendix A. Each of these items was repeated 4 times in the series for a total of 96 items. The last set of stimuli constituted the probes. These were 6 items directly relevant to the scenario tested by the sequence. The six probes for each scenario are also listed in Appendix A. Each probe was repeated 4 times, so that there were 24 probe trials. Note that for the innocent the series consists of targets and irrelevants, with 17% of the former and 83% of the latter. For the guilty, 17% of the trials are targets, 17% are probes, and 66% are irrelevants. The design of the test is summarized in Table 1.

Note that the subject pressed a switch in response to every stimulus. One hand was used to respond to the targets and the other hand was used to respond to the probes and the irrelevant stimuli. Like probes, targets were relatively rare, appearing once in every six stimuli.

Prior to each block, a list of the target stimuli for that block appeared on the screen. The experimenter read the list aloud, then the subject read the list aloud, and then the subject was asked to recall the list and was corrected if any errors or omissions occurred. The subject was instructed to press one microswitch following the presentation of a target stimulus, and another microswitch following any other stimulus. Subjects were instructed to press the switch as quickly and accurately as possible. The list of target stimuli was erased from the screen before the stimuli were presented.

Every 36 trials (that is, following one presentation of each stimulus) the stimuli were randomized again, and the next 36 trials were presented. This was repeated four times each block, for a total of 144 trials.

[2]The tests conducted with the conventional polygraph will not be discussed in this report. Using conventional polygraphic techniques, we were not able to detect the subjects' guilty knowledge. However, the circumstances of testing were quite different from those used by professional polygraphers, hence we do not believe any conclusions can be drawn from this phase of our study. We do note, however, that the conventional polygraph was not used *concurrently* with the ERP recording. Rather, a GKT patterned after Lykken (1981) was administered subsequent to the ERP test.

Table 1
Types of stimuli and predicted ERPs

Stimulus Type	Relative Frequency	Description	Instructions	Stimulus Evaluation	Predicted ERP
Target	1/6	Relevant to task (not to crime)	Right button press	Rare, Relevant	P300
Irrelevant (frequent)	2/3	Irrelevant to task and crime	Left button press	Frequent Irrelevant	No P300
Probe	1/6	Relevant to crime (not to task)	Left button press (treat like irrelevant stimuli)	If innocent: Frequent, Irrelevant (indistinguishable from irrelevant stimuli)	No P300
				If guilty; Rare, Relevant	P300

We required the subjects to press buttons to ensure that they would actually pay attention to each stimulus and perform the stimulus classification that is a prerequisite for the elicitation of the P300. However, the response times to the different types of stimuli may differ for guilty and innocent individuals, due to the increased task difficulty facing a guilty subject. An innocent subject can simply press one button for stimuli that are familiar, and the other button for all other stimuli. A guilty subject must distinguish among three types of stimuli, two of which are familiar. He must respond differently to stimuli that are familiar because of the instructions regarding responses to a subset of the stimuli (targets) and stimuli that are familiar because of the scenario he has enacted (probes). In addition to making the probes particularly salient and contributing to the probe P300 amplitude, this may lead to slower response times for probes than irrelevants in a guilty subject. However, because reaction time can easily be voluntarily manipulated, it is not suitable as a measure of guilt or innocence. In particular, the lack of a slower reaction time to probes may easily be produced by a shift in strategy, and thus is not an indication of innocence.

Data Acquisition

The electroencephalogram (EEG) was recorded using Ag/AgCl Beckman Biopotential electrodes placed at the F_Z (frontal), C_Z (central), and P_Z (parietal) sites (10-20 International system), and the right mastoid. All sites were referred to the left mastoid. In off-line analysis, half of the right mastoid/left mastoid signal was subtracted from each channel, so that the reference was in effect the average of the mastoids. Electro-oculogram (EOG) was recorded from sub- and supraorbital electrodes (above and below the right eye). The subjects were grounded at the forehead. Electrode impedance did not exceed 5 Kohm. Brain electrical activity was amplified by Van Gogh amplifiers with low- and high-pass filters set at half-amplitude frequencies of 35 and 0.02 Hz, respectively. These signals were digitized at a rate of 100 samples per second. ERPs and reaction times were recorded on tape for off-line analysis.

Prior to data analysis, all data were digitally filtered using a 49-point, equal-ripple, zero-phase-shift, optimal finite impulse response low-pass filter with a passband cutoff frequency of 6 Hz and a stopband cutoff frequency of 8 Hz. (For a discussion of digital filtering of ERPs, see Farwell, Martinerie, Bashore, & Rapp, 1991.)

All trials, including those with the EOG artifact, were recorded, and data from all trials were included in the reaction time results. However, only those trials with a range of EOG activity of less than 97.7 μV were included in the ERP analysis and in the trial counts that determined the number of trials presented.

Results

Event-Related Brain Potentials

The average ERP responses for artifact-free trials of each trial type at the P_Z electrode site for each of the 20 subjects in the guilty condition are displayed in Figure 1a. ERPs for the same subjects in the innocent condition are displayed in Figure 1b. The responses were as predicted. As can be seen in the figure, a large P300 was elicited by the target stimuli, but not by the irrelevant stimuli. The probes elicited a P300 in most subjects when they were relevant to the subject's "crime." A very small P300, if any, was elicited by the probes when the subject was "innocent."

Data Analysis

Our task in this study is to assess the similarity, for each subject, between the probe ERP and the ERP elicited by the other two stimuli. Furthermore, it was necessary to employ a method of analysis that would give a statistical confidence for each individual determination of guilt or innocence. However, in order to increase the signal-to-noise ratio to a workable level, it was necessary to collapse all of the trials of each type for an individual case to one average—and thus to eliminate any information we had on the distribution of ERP responses within an individual case. Moreover, any parametric estimate of the moments or distribution of correlations would not be valid, because the distribution of correlations violates the assumption of normality.

The statistical technique of bootstrapping (Efron, 1979; Wasserman & Bockenholt, 1989) provides one solution to this problem. To evaluate the significance of the apparent differences in Figure 1, we compared the three trial types using an iterative sampling bootstrapping procedure. Bootstrapping provides an estimate of the sampling distribution of a parameter when only a limited number of samples are available by obtaining many random subsamples from the available data and computing the parameter afresh for each of these subsamples. The distribution of these values approximates the actual distribution.

We used bootstrapping to estimate the sampling distribution of two correlations: the correlation between the average of the probe trials and the average of the irrelevant trials, and the correlation between the probe average and the target average. In our computations we used "double-centered" correlations (i.e., the grand mean for all trials of all types was subtracted from the probe, target, and irrelevant average waveforms prior to the correlation computations). If the correlation between the probe and target trials is significantly greater than the correlation between the probe and irrelevant trials, then we can conclude that the probe ERP responses are more similar to the target ERP responses (in which a P300 is present) than to the irrelevant ERP

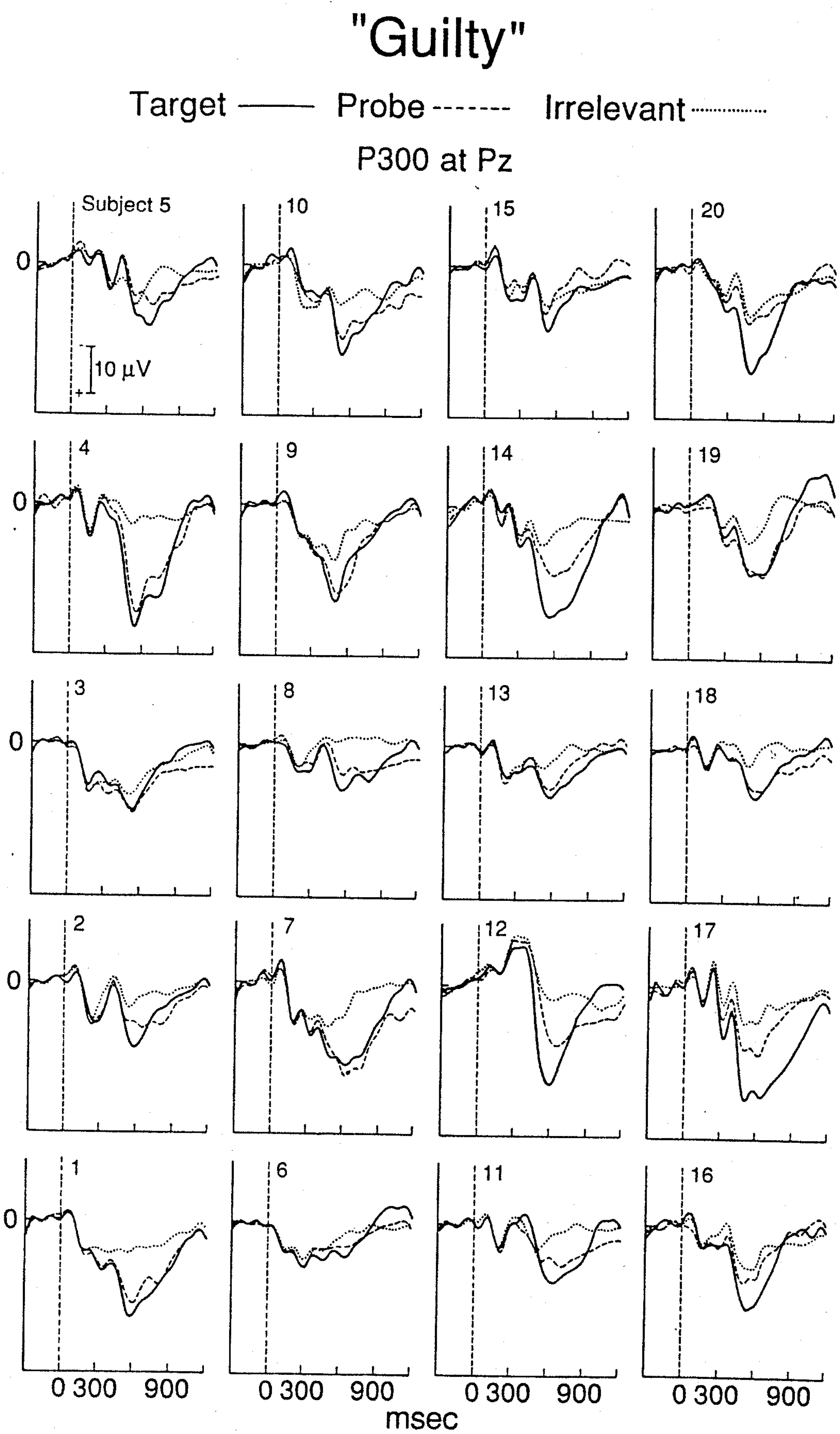

Figure 1a. ERPs for each of 20 subjects in the "guilty" condition (Experiment 1). Note that the probe waveform is clearly distinguishable from the irrelevant waveform, and similar to the target waveform.

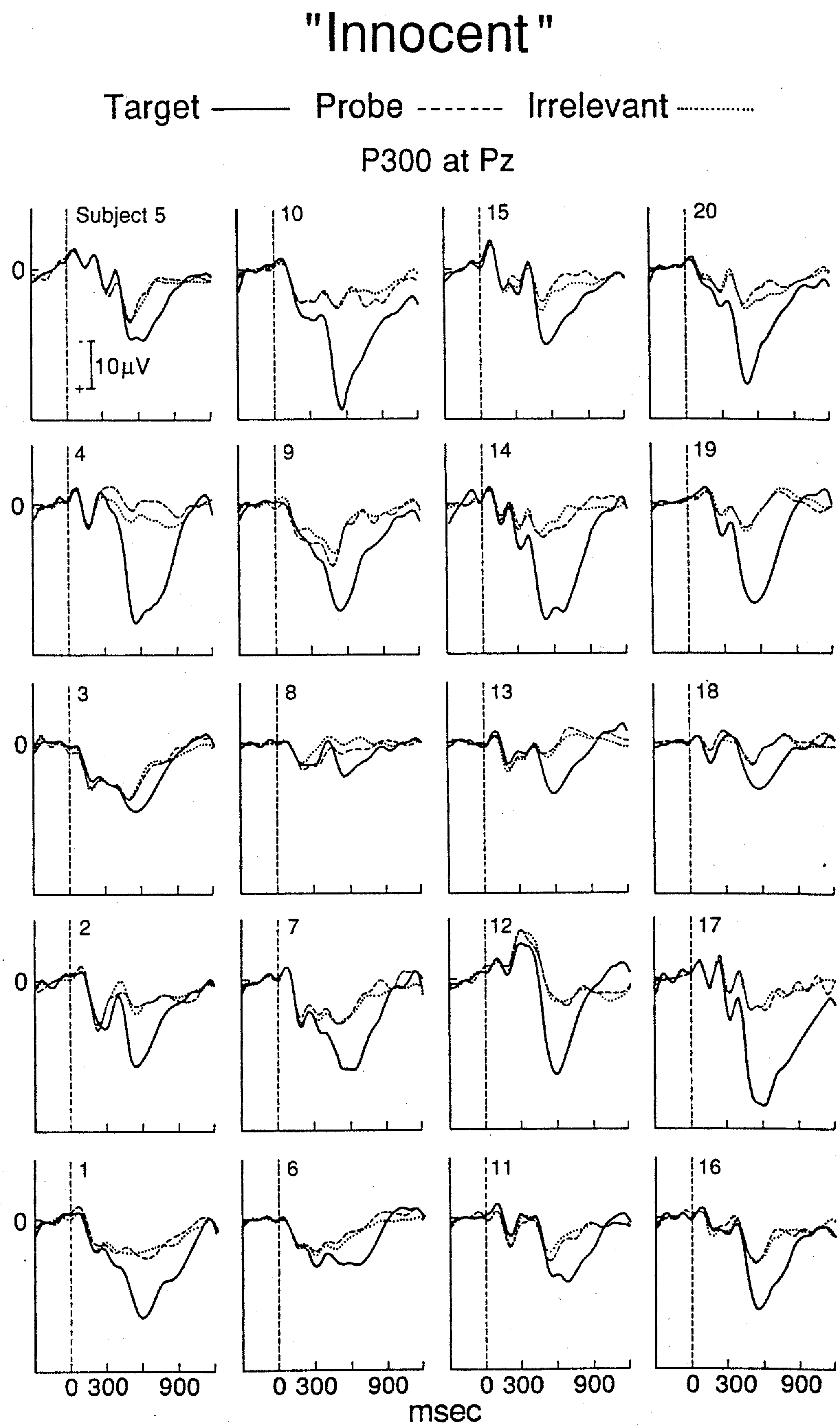

Figure 1b. ERPs for each of 20 subjects in the "innocent" condition (Experiment 1). (Subjects are the same as in Figure 1a.) Note that the probe waveform is similar to the irrelevant waveform.

responses (in which there is no P300). If this is the case, then we can conclude that the subject recognizes the probes as a separate, rare category—that is, of crime-relevant events—and therefore that the subject is "guilty." Similarly, if the correlation between the probe and irrelevant trials is greater than the correlation between the probe and target trials, then we can conclude that the subject is "innocent."

The procedure was as follows. We averaged each four irrelevant trials, so we had an equal number of probe, target, and irrelevant trials, approximately 72 of each (24 from each of three blocks). We created 100 random samples. The samples were taken with replacement. In each of 100 iterations we selected 72 of each of the three types of trials, yielding for each iteration three average ERPs. Each average was based on the 72 epochs selected for that iteration from one of the three trial types. We computed the probe-target and probe-irrelevant correlations for each iteration. Thus, the process yielded two groups of 100 correlations each. The distribution of these 100 correlations served as an estimate of the sampling distributions of the correlations. We then compared the distributions of the probe-irrelevant and probe-target correlations.

For each subject we counted the number of iterations on which the probe/target correlation exceeded the probe/irrelevant correlation. This value is called the "bootstrap index" in the following discussion. In Figure 2 we show the distribution of the bootstrap index for the 40 tests we conducted. The number of tests, labeled for guilt and innocence, corresponding to each index is plotted in the figure. It is evident that most guilty tests are associated with the lower values of the bootstrap index, whereas the higher values of the index are associated with tests in which the subjects did not possess the concealed knowledge. Five tests, two of the guilty and three of the innocent, fall in the middle of the range.

A decision regarding the guilt or innocence of a given subject depends on the extent to which the bootstrap index exceeds a criterion. Thus for example, we can decide to require that at least 90% of the iterations will declare the subject as guilty before guilt is accepted (bootstrap index of 0.10). A corresponding low limit on the index can be set, which if passed, the subject will be declared innocent. The data in Figure 2 suggest that we have considerable leeway in setting the criteria. This point is made also in Figure 3, which plots the outcomes of all possible decision rules for all the tests. It is clear that any "guilty" criterion that is greater than 0.06 and less than 0.36 will correctly identify 18 of the 20 subjects as guilty and will not misclassify any innocents as guilty. If we set the innocent criterion at any value greater than 0.47 and less than 0.80, 17 of the innocent subjects and none of the guilty subjects will be considered innocent. If we declare all subjects whose index falls between these two limits "indeterminate," we will find that 5 of the tests lead to indeterminate results whereas the remaining 35 tests lead to correct classification of the subjects. In no case, given the indeterminate class, do we have either a false positive or a false negative.

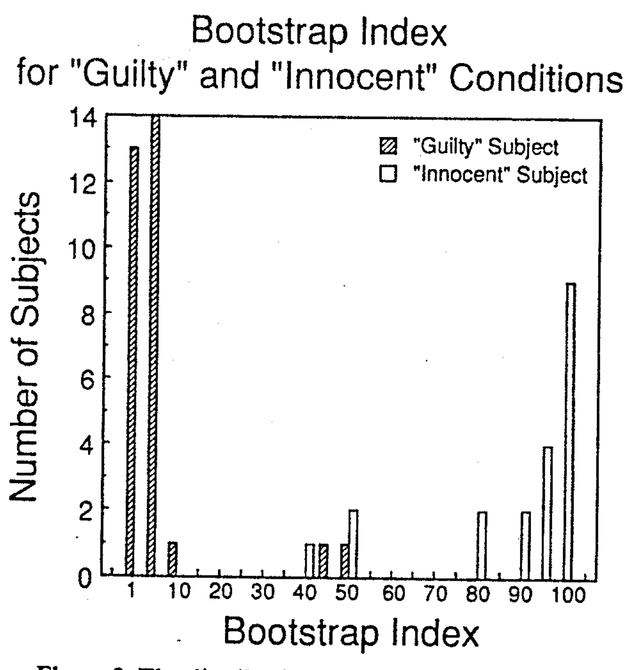

Figure 2. The distribution of the bootstrap statistic for all 40 tests conducted in Experiment 1. Dark bars indicate the number of subjects who were "guilty" and were assigned a given bootstrap value. Light bars show the same data for the "innocent" subjects.

The results of the bootstrapping analysis for one set of criteria are tabulated in Table 2. Table 2A summarizes the accuracy of determinations. Again, we have considerable leeway in setting the criteria while maintaining high accuracy of classification. Except for these 5 subjects, whose results are neither strongly "innocent" nor strongly "guilty," all of the guilty subjects have scores of .06 or less and all of the innocent subjects have scores of .80 or more. Tables 2B and 2C list the determinations for each of the guilty and innocent cases respectively. These tables also tabulate the bootstrap index, the proportions of iterations of the bootstrap procedure in which the probe-target correlation was greater than the probe-irrelevant correlation (i.e., the statistical confidence for each determination). Note that a high bootstrap index is an indication of "innocence" and a low bootstrap index is an indication of "guilt."

Scalp Distribution

In addition to the data for the parietal (P_Z) electrode site illustrated in Figure 1, we recorded data

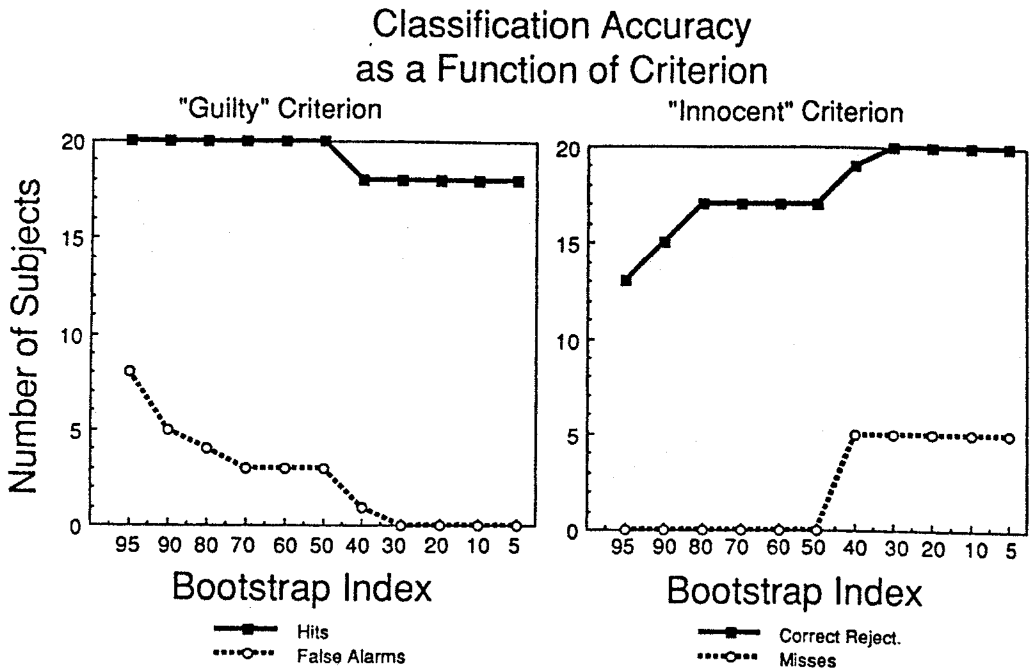

Figure 3. Accuracy of the guilty/innocent classification as a function of the bootstrap statistic used for determining guilt or innocence. A "hit" is a guilty subject classified as guilty; a "correct rejection" is an innocent subject classified as innocent; a "false alarm" is an innocent subject classified as guilty; and a "miss" is a guilty subject classified as innocent.

Table 2

2A: ACCURACY OF DETERMINATIONS

Decision	Subject State Guilty	Innocent	Total
Guilty	18	0	18
Innocent	0	17	17
Indeterminate	2	3	5
Total	20	20	40

Predictive Values	
Positive	Negative
100%	100%

Validity (excluding inconclusives)	100%
Validity (including inconclusives)	87.5%

Table 2A: Accuracy of determinations. In the 87.5% of the cases where a determination was made, 100% of the determinations were accurate. Positive and negative predictive values reflect the probability that guilty and innocent subjects, respectively, will be correctly determined, when a determination is made (i.e., excluding indeterminates). Validity reflects the overall probability of correctly determining a subject's state.

Decision rule:

Bootstrap statistic < .10 = = > Guilty
Bootstrap statistic < .70 = = > Innocent
Bootstrap statistic > = .10 and = < .70 = => Indeterminate

from the midline frontal (F_Z) and central (C_Z) sites. Subjects exhibited the usual parieto-central scalp distribution for the P300 in those conditions in which a P300 was present (i.e., in response to target stimuli in both the "innocent" and "guilty" conditions and to probe stimuli in the "guilty" condition). In most subjects, P300s showed maximum amplitude at P_Z, with a simultaneous smaller positive deflection at C_Z, and a still smaller one at F_Z. Three out of 20 subjects exhibited a C_Z-maximal P300, with a slightly smaller positive deflection at P_Z than at C_Z. Such a distribution across subjects of P300 scalp distributions is typical (Fabiani et al., 1987).

We performed the bootstrapping procedure using data from all three electrode sites in order to see whether or not the additional information provided by scalp distribution could contribute to increased accuracy of determinations of "innocence" or "guilt." We found that, due to greater variability in P300 amplitude and shape at F_Z and C_Z than at P_Z, including these additional channels in our bootstrapping analysis did not improve our ability to make accurate determinations. Collecting and visually inspecting the F_Z and C_Z data, however, did serve a useful purpose: Finding a scalp distribution typical of P300 made it more clear that the component we quantified at P_Z was indeed the P300. It is possible that some of the discriminating power was contributed by components other than the P300. From a practical viewpoint this is not an issue. As long as the system yields a valid decision regarding the subject's standing, the percent of the discriminating power contributed by different components is not a major issue at this stage of the development of the technique.

Overt Response Measures

Mean button-press response times for all trials (including EOG-contaminated trials) for each sub-

Table 2 B & C

2B: GUILTY CONDITION		
Subject #	Determination	Statistical Confidence
1	Guilty	.00
2	Guilty	.01
3	Guilty	.02
4	Guilty	.00
5	Guilty	.00
6	Guilty	.06
7	Guilty	.00
8	Guilty	.00
9	Guilty	.00
10	Guilty	.00
11	Guilty	.00
12	Guilty	.03
13	Guilty	.01
14	Guilty	.00
15	Guilty	.02
16	Indeterminate	.45
17	Guilty	.03
18	Guilty	.00
19	Guilty	.00
20	Indeterminate	.43

2C: INNOCENT CONDITION		
Subject #	Determination	Statistical Confidence
1	Innocent	.80
2	Innocent	.95
3	Indeterminate	.47
4	Innocent	1.00
5	Innocent	1.00
6	Innocent	.98
7	Innocent	.93
8	Indeterminate	.36
9	Innocent	1.00
10	Innocent	.99
11	Innocent	.96
12	Innocent	.91
13	Innocent	1.00
14	Innocent	.83
15	Innocent	1.00
16	Innocent	.96
17	Innocent	1.00
18	Indeterminate	.46
19	Innocent	1.00
20	Innocent	1.00

Tables 2B & 2C: Determinations and statistical confidence. Bootstrap statistic is the proportion of iterations (out of 100) where the correlation between the probe and irrelevant waveforms at P_Z was greater than the correlation between the probe and target waveforms. Note that a higher value indicates "innocence" and a lower value indicates "guilt."

ject in each condition are presented in Table 3. The response times are as predicted. Probe response times tend to be slower than irrelevant response times in the guilty condition, but not in the innocent condition. Also, when a given subject is guilty, the button presses in response to target stimuli also tend to be slower and less accurate than when the same subject is innocent. However, as mentioned above, because reaction time may be easily manipulated, it is not suitable as a measure of the presence of knowledge.

Table 3

Mean reaction times to the probe, irrelevant, and target stimuli for each subject in each condition

	Reaction Times (ms)		
Subject #	Target	Probe	Irrelevant
Innocent Condition			
1	918	775	775
2	1001	927	933
3	919	792	775
4	900	750	765
5	839	746	751
6	838	759	744
7	948	820	853
8	982	829	776
9	887	749	768
10	982	805	806
11	904	736	744
12	951	839	838
13	892	756	780
14	855	747	744
15	928	742	745
16	915	813	827
17	874	744	743
18	948	816	813
19	887	825	832
20	861	764	776
Averages	911	786	789
Guilty Condition			
1	941	880	810
2	1061	1087	936
3	999	1031	880
4	937	877	762
5	863	784	730
6	906	808	762
7	965	911	835
8	906	847	773
9	1017	872	781
10	1146	1111	905
11	950	774	759
12	972	890	826
13	1033	991	842
14	871	767	739
15	977	869	779
16	944	869	831
17	920	792	769
18	965	833	792
19	927	933	850
20	857	785	774
Averages	957	885	806

Discussion

The results confirm our prediction that the P300 can be used to identify those subjects who were familiar with the tested scenario. Inspection of the averages obtained from each subject indicates that the predicted pattern was obtained from virtually all subjects. Yet, it is important to avoid reliance on the gross waveforms when decisions are made with regard to individuals, decisions that may have serious consequences for the individual. We believe that such decisions should take into account the inherent variability of the data. The decision rules we base on the bootstrapping procedures adopt a conservative approach. It is gratifying to note that in no case did the bootstrapping analysis lead to an erroneous decision. That is, we were led neither to false positives nor to false negatives by the analysis. Instead, the analysis recognized that it did not have adequate information in 12.5% of the cases.

The results of this successful test may be viewed with some skepticism by those who believe that laboratory experiments using mock crimes do not provide an adequate test of an interrogative procedure. We agree that implementation of the ideas embodied in our procedure require extensive testing in "realistic" settings. We can, however, offer here a test conducted in an admittedly nonstressful setting, which did, nevertheless, interrogate subjects about "real life" transgressions of which they were definitely guilty. This data set is also of interest because it examines the efficacy of the P300-based Guilty Knowledge Test in circumstances in which the concealed knowledge derives from incidents that occurred at intervals ranging from weeks to months before the P300 test was conducted. Although it is true that the subjects' memory of the incidents was refreshed by the discussion of the incidents when they were recruited, the results do extend the scope of this feasibility test.

EXPERIMENT 2

The results of Experiment 1 clearly show the effectiveness of this paradigm in detecting guilty knowledge regarding a mock crime. The purpose of Experiment 2[3] was to examine the feasibility of the system in detecting guilty knowledge regarding actual crimes, which were not committed as a part of a laboratory study and may have taken place a considerable time prior to the testing situation. We tested 4 undergraduates who admitted having participated in minor crimes or socially undesirable activities (e.g., being arrested for underage drinking).

Method

Subjects

Four undergraduates at the University of Illinois served as subjects. The students were recruited by our advertising (through word of mouth) for subjects who had committed minor crimes or transgressions. The subjects were responsible for four transgressions, each being guilty of one and innocent of the other three.

Procedure

The experimental design was essentially the same as for Experiment 1, except for the modifications described below. The stimuli were presented visually, each stimulus a two-word phrase ranging from two to six syllables total. The probe stimuli were items relevant to the crime in question (e.g., the place where the crime took place or the name of another person involved). For each of the six probe stimuli there were one target and four irrelevant stimuli, as in Experiment 1. The target and irrelevant stimuli corresponding to each of the probe stimuli were items of the same type (e.g., a location where the crime did not take place, a fictitious name). Thus, the probe and irrelevant items were indistinguishable except to the guilty person. There was no training session or mock crime, because the test focused on an actual crime that had already taken place. Instead of pressing a button in response to target items as in Experiment 1, subjects were instructed to count the target items and to ignore the probe and irrelevant items. Subjects were asked for their tally at the end of each block of trials.[4] The target items were displayed at the bottom of the CRT throughout each block of the testing session as a memory aid.

Each subject was tested on his or her own crime ("guilty" condition), and also on another crime about which he or she knew nothing ("innocent" condition). The stimuli for the innocent condition for each subject consisted of the stimuli relevant to another subject's crime.

Results

The waveforms for each of the subjects in the "innocent" and "guilty" conditions are presented in Figure 4. Results are as predicted. It can be seen from the figure that there is a large P300 in response

---

[3]Experiment 2 was, in fact, our initial attempt to validate the concept embedded in the P300-based Interrogative Polygraphy method described in this paper. It was the success of our procedure in the context of Experiment 2 that led to our undertaking the large scale validation project described here as Experiment 1.

[4]Note that this counting task may not be as effective as the button press task in ensuring that subjects actually attend to and classify each stimulus. The button press was an innovation that was introduced after these data had been collected (see Farwell & Donchin, 1986, 1988, 1989).

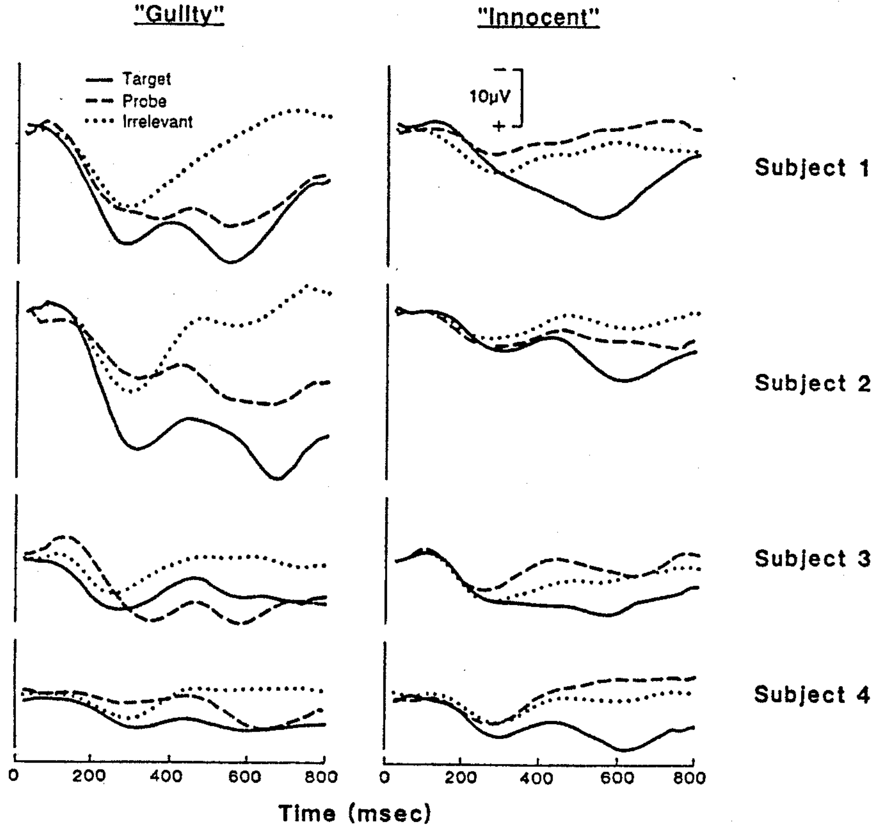

Figure 4. ERPs for 4 subjects in the "innocent" and "guilty" conditions (Experiment 2).

to the targets for all subjects in both conditions, and a very small P300, if any, in response to the irrelevants. In the "guilty" condition, all subjects show a P300 to the probe stimuli similar to their P300 to the targets. In the "innocent" condition the probe responses are similar to the irrelevants, and do not contain a large P300.

As in Experiment 1, we employed bootstrapping to quantify the differences that can be seen in Figure 4. The results are displayed in Table 4.

As in Experiment 1, the system proved highly reliable in distinguishing between the presence and the absence of guilty knowledge. The accuracy of determinations was the same in Experiment 2 as in Experiment 1: 100% correct in the cases in which a determination was made, with 12.5% indeterminate. All of the determinations, both "innocent" and "guilty," were made with a very high statistical confidence.

GENERAL DISCUSSION

The studies reported here were designed as a test of the feasibility of using event-related brain potentials (ERPs) in Interrogative Polygraphy (commonly referred to as "lie detection"). The data confirm the feasibility of designing a guilty knowledge test using the amplitude of the P300 component of the ERP as the measure used to detect the guilty knowledge. It is important to examine the logic of guilty knowledge tests so that the implications of our results for Interrogative Polygraphy can be placed in perspective. The conventional ANS-based Guilty Knowledge Test (GKT) was developed by Lykken (1959, 1960) as a procedure that was designed to circumvent the ambiguities associated with the control question technique. Lykken's procedure presents the subject with explicit questions that are directed to the crime being investigated. The test is a multiple choice test and the subject is provided with several equally plausible answers one of which is the crime-relevant test item.

The key assumption of the GKT is that there is some information about the episode that is known only to the investigator and to people who did participate in the episode. This information is the "guilty knowledge." Any implementation of the GKT sets up conditions in which the subject is presented with a sequence of stimuli, among which are stimuli that are distinct from all other stimuli by virtue of the guilty knowledge. The tests work if the distinctiveness of the items reflecting guilty knowledge is associated with a differential response in the bodily system whose activity is being monitored by the polygrapher.

As Furedy (1986) reminds us, the GKT does not detect deception per se. That is, the technique does

Table 4

4A: ACCURACY OF DETERMINATIONS

Decision	Subject State Guilty	Innocent	Total
Guilty	4	0	4
Innocent	0	3	3
Indeterminate	0	1	1
Total	4	4	8

Predictive Values	
Positive	Negative
100%	100%

Validity (excluding inconclusives)	100%
Validity (including inconclusives)	87.5%

Table 4A: Accuracy of determinations. In the 87.5% of the cases where a determination was made, 100% of the determinations were accurate. Positive and negative predictive values reflect the probability that guilty and innocent subjects, respectively, will be correctly determined, when a determination is made (i.e., excluding indeterminates). Validity reflects the overall probability of correctly determining a subject's state.

Decision rule:

Bootstrap statistic < .10 = = > Guilty
Bootstrap statistic > .70 = = > Innocent
Bootstrap statistic > = .10 and = < .70 = => Indeterminate

4B: GUILTY CONDITION

Subject #	Determination	Statistical Confidence
1	Guilty	.07
2	Guilty	.03
3	Guilty	.01
4	Guilty	.02

4C: INNOCENT CONDITION

Subject #	Determination	Statistical Confidence
1	Innocent	.99
2	Indeterminate	.27
3	Innocent	.96
4	Innocent	1.00

Tables 4B & 4C: Determinations and statistical confidence. Bootstrap statistic is the proportion of iterations (out of 100) where the correlation between the probe and irrelevant waveforms at P_z was greater than the correlation between the probe and target waveforms. Note that a higher value indicates "innocence" and a lower value indicates "guilt."

not directly assess the truth value of the subject's assertion. In that, it is similar to all other methods of Interrogative Polygraphy. Instead, as Lykken (1974) points out, the "basic assumption of the guilty knowledge test is that the guilty subject will show stronger autonomic response to what he recognizes as the significant alternative than he would have shown without such guilty knowledge" (pp. 727–728). Lykken goes on to explicitly attribute the distinctive response to an "orienting reflex" elicited by the "correct" alternative. (For a detailed discussion of the relationship between P300 and the Orienting Reflex, see Donchin & Fabiani, in press; Donchin & Heffley, 1978; and Siddle & Packer, in press. A comprehensive review of ANS approaches to Interrogative Polygraphy is provided by Furedy, 1986.)

The thesis argued in this report is that it is possible to employ the logic of the GKT without relying on the activation of autonomic responses. The system presented here does not rely on the elicitation of such responses by these "correct alternatives," to use Lykken's terminology, but rather takes advantage of the fact that these alternatives would be the only stimuli used in the test whose presentation will activate a cognitive processing subsystem revealed by the appearance of the P300.

The P300, recorded within the oddball paradigm, is an obvious candidate for implementing a GKT, because in this paradigm the subject is called upon to discriminate between two categories into which the individual members of the series of stimuli can be classified. When one of the categories is rare its members will elicit the P300. The actual application of this paradigm in Interrogative Polygraphy presents a number of challenges (see, for example, Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987, and Rosenfeld, Angell, Johnson, & Qian, 1991).

The procedure reported by Farwell and Donchin (1986), which is examined in detail in the studies reported here, modified the oddball paradigm in a manner that provides the necessary control conditions and allows us to focus the subjects' attention on the stimuli without at the same time biasing them to be concerned with the crime relevance of the test items. We do so by assigning the subjects an arbitrary task which, although having nothing to do with the crime and the investigation, requires them to carefully monitor a series of events for the occurrence of targets. These targets have nothing to do with the issue at hand, yet they become relevant by virtue of the instructions the subjects receive. Because these targets occur rarely, they do elicit a markedly large P300 and thus provide a baseline measure of P300 amplitude that can be elicited from the particular individual, on the particular occasion. The amplitude and shape of these P300s serve as a yardstick against which the P300s elicited by the "correct alternatives," as Lykken called the "guilty knowledge" items, can be evaluated.

These correct alternatives are hidden among the irrelevant items, and for the innocent they are indiscriminable from these items. The subjects who possess the guilty knowledge distinguish these "probes" from the other irrelevant items. Hence, the elicitation of a P300 by these probes is *prima* *facie* evidence that the subject possesses the guilty knowledge. This procedure combines the virtues of two of the major forms of conventional Interrogative Polygraphy procedures, the Guilty Knowledge Test (GKT) and the Control Question Technique (CQT). As in conventional GKTs, we utilize crime-relevant stimuli which to the innocent are indistinguishable from the irrelevant stimuli. This provides for resistance to false positives. However, the conventional GKT provides no control items to which a response is virtually guaranteed to serve as a criterion for evaluating the response to relevant items. These control questions are needed to guard against an excessive number of false negatives (Office of Technology Assessment, 1983). This deficiency of conventional GKT is avoided here by introducing the target stimuli, which serve as control items.

These data have both specific and broad implications for Interrogative Polygraphy. It is evident that it is possible to devise procedures that utilize event-related brain potentials in the aid of interrogations. It remains to be seen whether this particular implementation of the GKT can be applied within field settings. There is a long-standing debate (see Furedy, 1986) in this area of study between those who believe that the effectiveness of Interrogative Polygraphy cannot be assessed in laboratory conditions and those who assume that the laboratory provides an effective test bed for such techniques. Those who doubt the value of laboratory tests tend to invoke the lack of genuine stress in the laboratory as the rationale for their skepticism. It seems, though, that this issue is one to be resolved by empirical investigation. The utility of P300 in real life settings can be determined only by a test conducted in such settings. Yet, we note that because the P300 is used here as an index of a cognitive rather than an affective activity, it is considerably less reasonable to discount laboratory demonstrations. Thus far there has been little evidence that P300 can be modulated by affective variables, except that the more relevant the stimuli are to the subject's task, and the more relevant the task is to the subject, the larger the P300 (see Johnson, 1986). It would seem, therefore, that as the overall significance of the test increases in real life interrogations, the technique's effectiveness will increase rather than decrease. It is necessary to emphasize that the demonstration reported in this paper does not constitute evidence that a P300-based GKT will work under all circumstances with the effectiveness achieved in our test. The conditions of the test were clearly very different from those found in actual investigations. Neither the motivation nor the level of involvement of our subjects was close to that experienced by real suspects. Much additional research is required if the concept we have outlined is to be implemented.[5]

The P300 is but one of several endogenous components of the event-related brain potential (ERP) that may be used in Interrogative Polygraphy. The vocabulary of ERP components is quite varied, and one assumes that the current list of well studied components is not exhaustive. Thus, there are a number of negative components (Hillyard, 1984; Näätänen & Picton, 1987) appearing within the first 150 ms after a stimulus, which are sensitive to changes in the directionality of the subjects' attention and to the occurrence of various mismatches between the expected and the obtained stimuli. Kutas and her coworkers (Kutas & Hillyard, 1980; Kutas & Van Petten, 1988) have described a component labeled N400, which is affected by the degree to which a word violates the linguistic constraints imposed by the context in which the word was presented. There are a number of event-preceding negativities, such as the Readiness Potential and the Contingent Negative Variation, which reflect both cognitive and motor preparatory processes (Rohrbaugh & Gaillard, 1983; Walter, Cooper, Aldridge, McCallum, & Winter, 1964).

Each of these ERP components can be marshalled in the service of appropriately designed Interrogative Polygraphy procedures. It is critical to understand, however, when designing such procedures it would be an unwise strategy to launch a search for what has often been called The Pinocchio Response. That is, it is very unlikely that any ERP component, or any feature of the EEG, will serve as a specific and unique indicator that the subject has lied, or was in any other way deceptive.

It seems prudent to assume that the psychophysiological measures will not, by themselves, provide the hoped-for specific and unique indicators of deception per se. A more likely strategy for the design of an Interrogative Polygraphy method begins with a comprehensive understanding of the psychophysiological foundations of the measures

[5]A reviewer of an early version of this paper has expressed a common perception among those who practice conventional polygraphy that the use of ERPs requires the use of "costly, cumbersome and complex equipment." This is simply not the case. A fully functional P300-based testing device can be implemented in a device not larger than a standard polygraph, and the attachments to the subjects need be no more complex than those used for measuring the skin conductance response, and certainly less annoying than the conventional blood pressure cuff.

one is planning to use. One must begin with an analysis of what is known of the antecedent conditions for a component (Donchin & Coles, 1988) and the functional significance of the component. These will provide clues regarding the processing subsystems that might be manifested by the component. This psychophysiological database needs to be interfaced with the interrogatory task. We conceive of the psychophysiological database as a core, around which the designer constructs a shell. The shell capitalizes on the nature of the component and structures a scenario in which stimuli are presented to the subject in such a manner that the ERPs they elicit can be interpreted unequivocally in the context of the interrogation.

The three-category oddball paradigm, which is the subject of this study, is an example of such a shell. It is one of many possible arrangements of stimuli that would elicit a P300. In other words, one can imagine numerous shells that attempt to place our knowledge of the P300 in the service of some application. However, not all shells are equally effective. Hiding the probes among the irrelevants yields a very effective shell. Using the crime-relevant items as the only rare category in an oddball paradigm is a very poor shell. Similar considerations apply to each of the ERP components enumerated above. Knowing their functional significance and antecedent conditions is necessary, but not sufficient, for ensuring effective designs for shells. The emergence of an ERP-based polygraphy will depend on the ingenuity with which shells are designed. Useful shells will emerge from a careful and methodical analysis of the properties of the different ERP components, not from a brute force search for deception indicators.

We conclude by noting that the results we report raise interesting questions with respect to the P300. Our success in detecting those subjects who were informed about the probes is somewhat puzzling. The probes were definitely not relevant to the task the subject was ostensibly performing. They were distinguished from the other irrelevant items solely by their association with the mock espionage scenario in which the subject participated. Why did such "irrelevant" items elicit a P300? We hypothesized, and demonstrated, that subjects are sensitive to such seemingly irrelevant items if these items are distinctive in some dimension that is important to the subject, even though it is not relevant to the task the subject is performing.

The implication is that subjects monitor events along dimensions other than those specified by the experimenter, and that when distinctiveness is detected across such irrelevant dimensions, it may trigger the processing subsystem manifested by the P300. It did appear plausible that crime-related items would play such a role, and our results indicate that they did. However, we cannot predict, in the general case, which dimensions along which items are distinct will, or will not, be noted by the subject when the dimension in question is formally irrelevant. This is a rather important question, because the concept of task relevance has played an important role in accounting for the P300 (Donchin et al., 1978; Johnson, 1986; Johnson & Donchin, 1978; Rösler, 1983). Frequently, task relevance is defined strictly in terms of the task assigned to the subject (e.g., Courchesne, Hillyard, & Galambos, 1975). This is not, however, a fully adequate definition, because subjects evidently extract information relevant to aspects of the situation that has little if anything to do with the assigned task. The present study did not examine in detail the range of distinctions that would play the role played by the probes in the present study. Such an analysis is clearly needed.

The interest in a detailed elucidation of the circumstances under which hidden probes will be effective in eliciting a P300 is important from the psychophysiological perspective, because of the contribution it will make to a better understanding of the P300 component. The procedure we used here for detecting whether subjects have information regarding an espionage mission (or a minor crime) can be extended to determine whether subjects make other distinctions of which they are unaware or for which they are not reliable witnesses.

REFERENCES

Barland, G.H., & Raskin, D.C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology, 12*, 321–330.

Ben-Shakhar, G., & Furedy, J.J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective.* New York: Springer-Verlag.

Courchesne, E., Hillyard, S.A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography & Clinical Neurophysiology, 39*, 131–143.

Donchin, E. (1981). Surprise!... Surprise? *Psychophysiology, 18*, 493–513.

Donchin, E., & Coles, M.G.H. (1988). Is the P300 component a manifestation of context updating? *The Behavioral and Brain Sciences, 11*, 355–425.

Donchin, E., & Fabiani, M. (in press). The use of event-related brain potentials in the study of memory: Is P300 a measure of event distinctiveness? In J.R. Jennings & M.G.H. Coles (Eds.), *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches.* Chichester, UK: John Wiley.

Donchin, E., & Heffley, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (pp. 215–217). EPA-6001-9-77-043, Washington, DC: U.S. Government Printing Office.

Donchin, E., Heffley, E., Hillyard, S.A., Loveless, N., Maltzman, I., Ohman, A., Rösler, F., Ruchkin, D., & Siddle, D. (1984). Cognition and event-related potentials. II. The orienting reflex and P300. *Brain and information: Event-related potentials. Annals of the New York Academy of Sciences, 425*, 39–57.

Donchin, E., Karis, D., Bashore, T.R., Coles, M.G.H., & Gratton, G. (1986). Cognitive psychophysiology and human information processing. In M.G.H. Coles, E. Donchin, & S.W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 244–267). New York: Guilford Press.

Donchin, E., Ritter, W., & McCallum, W.C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man* (pp. 349–441). New York: Academic Press.

Duncan-Johnson, C.C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology, 14*, 456–467.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1–26.

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, politics, and marriage.* New York: Norton.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). The definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P.K. Ackles, J.R. Jennings, & M.G.H. Coles (Eds.), *Advances in psychophysiology* (Vol. 2, pp. 1–78). Greenwich, CT: JAI Press, Inc.

Farwell, L.A., & Donchin, E. (1986). The "brain detector:" P300 in the detection of deception [Abstract]. *Psychophysiology, 24*, 434.

Farwell, L.A., & Donchin, E. (1988). The truth will out: Interrogative polygraphy with event-related brain potentials [Abstract]. *Psychophysiology, 25*, 445.

Farwell, L.A., & Donchin, E. (1989). Detection of guilty knowledge with event related potentials [Abstract]. *Psychophysiology, 26*(Suppl.), S8.

Farwell, L.A., Martinerie, J.M., Bashore, T.B., & Rapp, P.E. (1991). *Optimal digital filters for long latency event-related brain potentials.* Manuscript in preparation.

Furedy, J.J. (1986). Lie detection as psychophysiological differentiation: Some fine lines. In M.G.H. Coles, E. Donchin, & S.W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 683–701). New York: Guilford Press.

Hillyard, S.A. (1984). Event-related potentials and selective attention. In E. Donchin (Ed.), *Cognitive psychophysiology: Event-related potentials and the study of cognition. Vol. I: The Carmel Conferences* (pp. 51–72). Hillsdale, NJ: Erlbaum.

Hillyard, S.A., & Kutas, M. (1983). Electrophysiology of cognitive processing. In M.R. Rosenzweig & L.W. Porter (Eds.), *Annual review of psychology* (Vol. 34, pp. 33–61). Palo Alto, CA: Annual Reviews, Inc.

Johnson, R., Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*, 367–384.

Johnson, R., Jr., & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. *Electroencephalography & Clinical Neurophysiology, 44*, 424–437.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Van Petten, C. (1988). The N400 and language. In P.K. Ackles, J.R. Jennings, & M.G.H. Coles (Eds.), *Advances in psychophysiology* (Vol. 3, pp. 139–187). Greenwich, CT: JAI Press, Inc.

Lykken, D.T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43*, 385–388.

Lykken, D.T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology, 44*, 258–262.

Lykken, D.T. (1974). Psychology and the lie detector industry. *American Psychologist, 29*, 725–739.

Lykken, D.T. (1981). *A tremor in the blood: Uses and abuses of the lie detector.* New York: McGraw Hill.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*, 375–425.

Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation—A technical memorandum.* Washington, DC: U.S. Government Printing Office.

Podlesny, J.A., & Raskin, D.C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology, 15*, 344–358.

Raskin, D. C. (1986). The polygraph in 1986: Scientific, professional, and legal issues surrounding applications and acceptance of polygraph evidence. *The Utah Law Review, 1986*, 29–74.

Reid, J.E., & Inbau, F.E. (1977). *Truth and deception—The polygraph technique* (3rd ed.). Baltimore: The Williams and Wilkins Co.

Rohrbaugh, J.W., & Gaillard, A.W.K. (1983). Sensory and motor aspects of the contingent negative variation. In A.W.K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research: Endogenous components* (pp. 269–310). Amsterdam: North-Holland.

Rosenfeld, J.P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology, 28*, 319–335.

Rosenfeld, J.P., Nasman, V.T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: A new method of lie detection. *International Journal of Neuroscience, 34*, 125–129.

Rösler, F. (1983). Endogenous ERPs and cognition: Probes, prospects, and pitfalls in matching pieces of the mind-body problem. In A.W.K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research:*

Endogenous components (pp. 9-35). Amsterdam: Elsevier.

Siddle, D.A.T., & Packer, J.S. (in press). Memory and autonomic activity: The role of the orienting response. In J.R. Jennings & M.G.H. Coles (Eds.), *Handbook of cognitive psychophysiology: Central and autonomic nervous system approaches.* Chichester, UK: John Wiley.

Squires, K.C., Squires, N.K., & Hillyard, S.A. (1975). Decision-related cortical potentials during an auditory signal detection task with cued intervals. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 268–279.

Walter, W.G., Cooper, R., Aldridge, V.J., McCallum, W.C., & Winter, A.L. (1964). Contingent negative variation: An electrical sign of sensorimotor association and expectancy in the human brain. *Nature, 203*, 380–384.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology, 26*, 208–221.

Appendix A

The stimuli used in the present study

Scenario 1			*Scenario 2*		
Probe	**Target**	**Irrelevant**	**Probe**	**Target**	**Irrelevant**
Blue Coat	Green Hat	Brown Shoes Red Scarf Gray Pants Black Gloves	White Shirt	Green Tie	Beige Suit Red Vest Tan Belt Black Socks
Phil Jenks	Tim Howe	Ray Snell Neil Rand Gene Falk Ralph Croft	Dale Spence	Wayne Brant	Glenn Platt Walt Rusk Tod Ames Earl Dade
Op Cow	Op Pig	Op Horse Op Goat Op Sheep Op Mule	Op Spruce	Op Fir	Op Oak Op Birch Op Elm Op Pine
Rain File	Snow File	Hail File Wind File Sleet File Fog File	Owl File	Swan File	Wren File Duck File Crow File Goose File
Sub Plans	Ship Plans	Tank Plans Plane Plans Bomb Plans Gun Plans	Brass Plans	Steel Plans	Tin Plans Zinc Plans Lead Plans Iron Plans
Perch Street	Shark Street	Cod Street Carp Street Pike Street Trout Street	Lion Street	Fox Street	Deer Street Wolf Street Bear Street Elk Street

(Manuscript received May 24, 1990; accepted for publication October 20, 1990)