Are there alternatives to traditional polygraph procedures?

Article *in* Psychological Bulletin · January 1993

DOI: 10.1037/0033-2909.113.1.3

Psychological Bulletin
1993, Vol. 113, No. 1, 3-22

0033-2909/93/$3.00

Are There Alternatives to Traditional Polygraph Procedures?

Theodore R. Bashore and Paul E. Rapp

The most commonly used method for detecting deception is based on the assumption that lies given by a person in response to critical questions posed during a polygraph examination will elicit an identifiable pattern of autonomic reactivity. Critics of this method argue that a polygraph examination cannot detect lying because lying does not produce a distinct physiological response. They assert that the possession of information only the guilty person would be expected to have can be revealed in a polygraph examination, however, by the pattern of autonomic arousal presentation of this information elicits in a person who possesses it. In this article, the position is taken that the dependence of both procedures on autonomic measures diminishes their effectiveness and inhibits the development of alternatives. A few studies are reviewed that suggest that measures of brain electrical activity can be used to infer the possession of information in persons attempting to conceal it.

Traditional procedures for lie detection use multiple autonomic measures of physiological activation (e.g., electrodermal response, blood pressure, heart rate, and respiration) in conjunction with behavioral observations to infer truth or deception in the subject being examined. The assumption underlying these procedures is that the fear of being detected and punished induced in the subject by the examination process produces patterns of autonomic arousal and behavioral change associated with deception that can be identified by the person conducting the examination (Reid & Inbau, 1977). Advocates of these traditional procedures assert that the deceptive autonomic arousal pattern is apparent on polygraphic recordings taken during administration of a brief series of questions formulated by the examiner to differentially activate the subject's autonomic nervous system, that the behavioral patterns suggestive of culpability are also manifest during the various phases of the lie detection examination, and that they, too, can be identified by the examiner. The most commonly used procedure for lie detection in criminal investigations is the Control Question Technique (CQT), developed by John Reid in 1947 (Office of Technology Assessment [OTA], 1983; Reid & Inbau, 1977); the most commonly used procedure for lie detection in preemployment screening is the Relevant/Irrelevant Question Technique (RI; Lykken, 1981; OTA, 1983).

Critics of these traditional lie detection procedures have argued that the assumptions on which they are based are flawed and that there is essentially no sound scientific support for their validity (Ben-Shakhar, Lieblich, & Bar-Hillel, 1982, 1986; Furedy, 1987; Furedy & Heslegrave, 1988; Kleinmuntz & Szucko, 1982, 1984; Lykken, 1978, 1979, 1981, 1984; Saxe, Dougherty, & Cross, 1985; also see reviews in OTA, 1983, and Working Group of the Scientific Affairs Board [Working Group], 1986, reports).

Theodore R. Bashore, Department of Psychology, University of Northern Colorado; Paul E. Rapp, Department of Physiology, Medical College of Pennsylvania.

Correspondence concerning this article should be addressed to Theodore R. Bashore, Department of Psychology, University of Northern Colorado, Greeley, Colorado 80639.

These critics contend that the traditional procedures have not been subjected to rigorous scientific evaluation and that claims of the power of these techniques by their proponents are based on subjective and impressionistic observations. Moreover, the critics claim that the vast majority of reports on the efficacy of these procedures have been done by the advocates themselves, who presumably have a vested interest in promoting their trade. These critics also assert that among those trained in traditional lie detection procedures, there are very few behavioral scientists or psychophysiologists with advanced academic degrees who are competent to conduct the type of research needed to assess the validity of these procedures. Another, and fundamental, criticism made by opponents of the traditional lie detection procedures is that there is no physiological or behavioral response pattern indicative of lying, despite the claims made by its proponents. Indeed, in assessing the overall condition of research in traditional polygraphy, one critic has asserted the following:

> If proponents wish to convince the scientific community of the merits of polygraph lie detection, I submit that they will have to develop a more convincing case than the one currently on offer. Their case must be founded on studies which include the necessary controls for nonpolygraphic sources of information, that is, studies which compare the accuracy of assessments derived from case-file material and the subject's demeanour during questioning with that based on these sources plus the polygraphic record. I strongly suspect that such studies would confirm what the available data suggest: that polygraph lie detection adds nothing positive to conventional approaches to interrogation and assessment. (Carroll, 1988, p. 28)

For many of these critics, an alternative procedure, introduced by Lykken (1959), the Guilty Knowledge Test (GKT), is more desirable because it has a sounder theoretical basis. This test does not derive from the assumption that there is a distinct physiologic response, or a set of well-defined behavioral manifestations, associated with lying. Rather, its development was motivated by the assumption that autonomic arousal will be induced in the guilty subject by aspects of the transgressive act and situation known only to that person. Thus, the test was designed to identify guilt (as expressed in a cognitive process, recognition) as opposed to deception (as expressed in an emotional reaction, fear). Like the traditional procedures, the GKT has been developed for use with autonomic measures.

Criticism of traditional lie detection techniques and the assumptions on which they are grounded has given rise to an often vitriolic exchange between its critics, most notably David Lykken, and its proponents, most notably David Raskin. During the course of the debate by these two very highly respected psychophysiologists (dating to the 1970s; Lykken, 1978, 1979; Raskin, 1978; Raskin & Podlesny, 1979), use of traditional lie detection procedures has continued to be widespread and, in fact, has grown in criminal investigations and employment screening in the private and government sectors. Indeed, in 1983, the Reagan administration issued a national security directive, rescinded within a year, authorizing executive agencies and departments to require employees to take polygraph examinations during investigations of security violations. The Reagan administration's advocacy of the polygraph examination is consistent with a widely held public perception that it is infallible in the detection of deception. This is the case despite the growing number of published criticisms by behavioral scientists of the validity of these procedures. As is well known, lives of innocent people have been damaged, perhaps unalterably, by false positives on lie detection tests (e.g., see Kleinmuntz & Szucko, 1984; Lykken, 1981). The fact that two eminent scientists such as Lykken and Raskin can take such extreme views of the traditional lie detection techniques suggests that the measures on which they are based may be too elusive to provide indexes of processes as subtle and complex as guilt or deception.

In what follows, we review these procedures briefly (for a compendium of reviews on polygraphy, see Gale, 1988) and then describe work done with event-related brain potentials that is suggestive of a more direct, perhaps less elusive, procedure for assessing guilt (Farwell, 1991; Farwell & Donchin, 1986, 1988a, 1988b, 1991; Rosenfeld et al., 1988; Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987). We then suggest how this work can be extended to include a number of methodological and analytical refinements that may sharpen the discriminatory power of these brain electrical signals even further.

The Traditional Lie Detection Examination

The traditional approach to lie detection is exemplified in the procedures developed by John Reid and his associates (Reid & Inbau, 1977). These procedures are used both in preemployment screening and in criminal investigations but are used most commonly as part of a job-screening process, particularly if the position involves access to information that is deemed to be sensitive by the employer. The use of these procedures is often justified by an employer, however, as providing a tool by which the character of a prospective employee can be assessed. Under these circumstances, the test format that is typically used is the RI (Lykken, 1981; OTA, 1983; Reid & Inbau, 1977). In this test, the subject is asked a series of questions that are either irrelevant (e.g., "Did you graduate from high school?") or relevant (e.g., "Have you ever smoked marijuana?") to the issue being assessed. The assumption is that a person who has engaged in the undesired activities and denies having done so will be more reactive to the relevant questions than to the irrelevant questions. That is, it is assumed that the general untruthfulness of the person can be determined by the pattern of autonomic reactivity elicited in him or her by the relevant questions. To say the very least, this is a dubious assumption that has been the object of serious criticism (see reviews in Gale, 1988; OTA, 1983; Working Group, 1986), discussion of which extends beyond the scope of this article. Our discussion is limited, therefore, to the use of procedures that are most applicable to criminal investigations. In these circumstances, very precise pieces of information are available about the crime under investigation, thereby allowing questions in the interrogation to be tailored to specific details. Even under these conditions, however, there is a substantial amount of research that must be done before conclusions can be drawn about the sensitivity of either traditional or alternative procedures in differentiating the guilty from the innocent. There is an enormous leap from this type of application to one in which the assessment of character is attempted by using vague questions that refer to nonspecific events in one's past.

Central to the examination procedures developed by Reid and associates for use in criminal investigations is the CQT, given during the polygraph test phase of the lie detection examination. The Reid examination has several components, some of which are always given and others of which are given only if the results produced by the standard components are inconclusive. All subjects are given a pretest interview, lasting 30–90 min, during which the control questions are formulated (Reid & Inbau, 1977, suggest that 30–40 min is sufficient), an initial CQT polygraph test, a card test[1] on the polygraph, and a second CQT polygraph test. If the results of this sequence suggest that the subject is truthful, the examination is concluded. If the results suggest deception, the subject is interrogated by the examiner in an attempt to elicit an admission of guilt. Reid and Inbau report that about 25% of subjects fall into well-defined truthful or deceptive groups. The balance of subjects, some 75%, judged to be inconclusive after the second CQT, may be examined further with an assortment of other polygraphic tests (e.g., mixed question, yes, guilt complex, and peak of tension). Reid and Inbau assert that about 10% of the subjects remain inconclusive after this lengthier examination and are called back for reexamination.

The CQT is referred to as a specific-incident (or specific-issues) test that includes three types of questions concerning the matter under investigation: relevant, irrelevant, and control. These questions refer in a reasonably general fashion to the incident being investigated. Relevant questions pertain to the

[1] This test, also known as the *stimulation procedure* or *stim test*, is designed to convince the subject of the validity of the polygraph examination as a lie detector. Typically, the subject is asked to pick a playing card from a deck of cards, to take note of it, and to return it to the deck. The polygrapher then calls out the names of several cards, and the subject is instructed to respond with no to each name. As the subject does so, the examiner pretends to evaluate the polygraphic response. After the last response is made, the polygrapher correctly informs the subject of the card he or she selected and asserts that the machine revealed the choice. Unbeknownst to the subject, the deck of cards contains only one type of card.

critical incident (e.g., "Did you rob Fidelity Bank on the night of April 3?"), whereas irrelevant questions are unrelated to the incident (e.g., "Did you graduate from high school?"). Control questions are designed to be unrelated to the specific incident but nonetheless emotionally provocative for innocent subjects (e.g., "Have you ever stolen anything?"). They provide the critical questions for determining truth or deception. It is assumed that the skilled examiner will be able to formulate control questions that produce different responses in the innocent and guilty subject. The control question should elicit an emotional response in the innocent subject that exceeds the response to the relevant question, but the relevant question should elicit the larger emotional response in the guilty subject.

There are at least five elements of this examination that are of particular concern to its critics: (a) the reliance on behavioral observations by the examiner in arriving at a diagnosis of truth or deception (i.e., the claim that there is a constellation of behaviors that reveal honesty or deception in the subject), (b) the assumption that there is a physiologic response that is specific to lying, (c) the assumption of differential autonomic responsivity to the control questions by innocent and guilty subjects (i.e., the concept of the control question), (d) the difficulty in formulating appropriately sensitive control questions, and (e) the fact that scoring of the polygraph results is done by the examiners themselves and is largely subjective.

The common view of these critics is that the examination places an enormous burden on the examiner, who usually has very little training in the areas being evaluated, to make a very sensitive clinical psychophysiological judgment of the type few, if any, highly trained clinical psychologists or psychophysiologists could make (Furedy, 1987; Furedy, Davis, & Gurevich, 1988; Furedy & Heslegrave, 1988; Kleinmuntz & Szucko, 1984; Lykken, 1981; Szucko & Kleinmuntz, 1981). Indeed, Furedy and Heslegrave suggest that the examiner is called on to make a series of complicated and extremely subjective clinical, psychophysiological, and social psychological decisions over the extent of the examination to arrive at a diagnosis. The subjectivity integral to the examination process, critics also assert, precludes standardizing its procedures and promotes idiosyncratic approaches to diagnosis among polygraphers. Thus, to its critics, the traditional lie detection examination is actually a complex psychological examination that includes a psychological test, the polygraph test. However, the polygraph test neither meets the required criteria of standardization nor established criteria of reliability and validity for psychological tests (Kleinmuntz & Szucko, 1984; Lykken, 1981; Szucko & Kleinmuntz, 1981). These views are nicely summarized by Furedy and Heslegrave (1988):

> It is also important to recognize the important role played by the attitudes of evaluators toward accepting errors, and weighing these errors depending upon the circumstances. Because the circumstances are known by the evaluator prior to the administration of the polygraph, and because the test is not standardized, it is likely that not only will the outcome be judged on the basis of examinee circumstance and examiner attitude, but also the administration of the test will be shaped by these prejudices. Because the test is psychological in the sense of involving a complex interview-like interaction between examiner and examinee, any biases in designing and administering the test are likely to produce outcomes that are consistent with those biases. So different individuals accused of different crimes may be given quite different tests, even though all of those tests are called by a single name—a polygraph test. Indeed, the term *test* itself is potentially misleading, because it suggests relatively standardized instruments such as IQ tests that, although controversial, give essentially the same results across competent operators. (p. 224)

As testimony to the severity of the demands placed on the examiner, witness Reid and Inbau's (1977) summary of an effective pretest interview and their description of how the optimal control questions are formulated:

> A properly conducted pretest interview has the effect of allaying the apprehensions of the truthful subject, nullifying the hostility of the hostile one, and stimulating the lethargic liar into displaying Polygraphic responsiveness. (p. 24) The examiner will seek to develop a *control question*—a question about an act of wrongdoing *of the same general nature as the main incident under investigation,* and one to which the subject, in all probability, will lie or to which his answer will be of dubious validity in his own mind. (p. 28) Whenever the subject says "no" to any one of the attempts at formulating a satisfactory control question, and, based upon the subject's behavioral symptoms, the examiner feels that the "no" answer is either false or that there may be some doubt in the subject's mind as to its accuracy (or that he will in any event harbor concern over it), then the examiner has achieved his objective of securing a control question. (p. 29)

Thus, the examiner must induce the appropriate psychological set in the subject in a 30–40 min pretest interview while at the same time formulating the critical control questions. Moreover, the examiner has the additional task of convincing the subject of the scientific credibility of the procedure (also attempted with the card test). To develop the control questions, the examiner is required to infer covert activity (i.e., what is happening in the subject's mind) from overt activity (i.e., behavior) and then to devise control questions that will be differentially sensitive to the inferred discomfort of the innocent person vis-à-vis the discomfort invoked by the relevant questions. This is a very sophisticated piece of psychological engineering (Lykken, 1981). Critics generally agree that it is naive to assume that control questions can be devised to accomplish this end. Moreover, these questions do not meet the scientific standard for a control; they are not similar to the relevant questions in all respects save their connection to the specific crime.

Scoring in the Reid lie detection examination is based on the examiner's integration of the behavioral and physiological information collected; scoring of the polygraph record is achieved through an experimental, intuitive process (referred to as *global*) that does not have any specific measurement criteria. Note that Podlesny and Raskin (1978) have reported that the behavioral observations do not contribute importantly to the final diagnosis when an alternative scoring procedure, numerical scoring, is used. This technique was devised by Backster (1962), and improved by Raskin (1976), in an attempt to reduce the subjective element involved in evaluating the polygraph results. It makes no assumption of a specific physiologic lie response. It is objective in the sense that differences between physiological responses to the relevant and control questions in each response channel are scored, from 3 to −3, and summed algebraically. The score is arrived at by subjective and qualitative means, however. In Barland and Raskin's (1975) terms, it is *semiobjective,* in that the examiner judges the degree of difference and assigns the score. Because polygraphers always score their own records, this method of scoring is open to bias factors that are well documented in the behavioral and physical sciences. Nonetheless, numerical scoring has been reported to be highly reliable and to produce near-perfect agreement among examiners in successfully identifying guilty subjects (Kircher & Raskin, 1988; Podlesny & Raskin, 1978; Raskin & Hare, 1978).

Note, however, that high interrater agreement and high success rates in identifying the guilty should not be confused with validity (Kleinmuntz & Szucko, 1984; Lykken, 1981, 1984; Saxe et al., 1985). There is also a high rate of false positives (innocent called guilty) often ignored by proponents of the traditional procedures. Indeed, with some notable exceptions (e.g., Barland & Raskin, 1975; Horvath, 1977), attention within the traditional polygraphic community has focused on the elimination of false negatives. This preoccupation has resulted in the acceptance of a high incidence of false-positive classifications in which innocent people are identified as guilty. These failures produce computed validities that exceed chance by a very small margin (see references below), a fact that is not generally known by the public or by the larger scientific community. A computer-based numerical scoring system was developed by Kircher and Raskin (1988) to provide objectivity, but this cannot overcome the problems of the CQT or of the functions being measured.

The Guilty Knowledge Test

The success of traditional lie detection procedures rests on a set of controversial assumptions about the behavioral and physiological manifestations of lying, the ability of the examiner to infer deception from these responses, and the power of the CQT to establish a physiologic distinction between honest and deceptive responses. It has been demonstrated in signal detection analyses of field polygraph studies that examiners may bias their diagnoses to reduce the number of guilty subjects diagnosed as innocent (i.e., false negatives) and, in so doing, may increase the number of innocent subjects misdiagnosed as guilty (i.e., false positives, e.g., Ben-Shakar et al., 1982; see also Szucko & Kleinmuntz, 1981, for a related analysis). Indeed, reviews of the literature have revealed false-positive rates for the CQT as high as 50%, with a mean in the mid-30% range (Kleinmuntz & Szucko, 1982, 1984; Lykken, 1981, 1984; Saxe et al., 1985; see also OTA, 1983, and Working Group, 1986, reviews). To circumvent these deficiencies, Lykken (1959, 1960, 1981) and others (Ben-Shakhar et al., 1986; Kleinmuntz & Szucko, 1982, 1984; Yohman, 1978) have advocated use of the Guilty Knowledge Test (GKT). Lykken (see review in Lykken, 1981) was motivated to develop this test on the basis of the following assumptions: (a) The task in a criminal investigation is to distinguish guilt from innocence, not to detect lying; (b) the guilty possess knowledge of the crime not available to the innocent; and (c) recognition of pieces of information unique to the crime will produce an arousal response in the guilty subject. Thus, this test is based on a cognitive link between important elements of the crime and the guilty subject; critical items are assumed to produce autonomic arousal in only those people who recognize them from the crime (i.e., the guilty).

In the GKT, the subject is given a series of multiple-choice questions, each of which has six choices, and is typically required to respond to each item simply by identifying it. The first choice, referred to as a *buffer,* is always incorrect. It is included to control for the fact that a large autonomic response is generally induced by the first stimulus in a sequence, irrespective of its significance. The five scored alternatives consist of one relevant and four control items. Control items in this test, in contrast to the control questions in the traditional polygraph test, actually do provide an index of how strong an innocent subject's autonomic response to the relevant item should be. Furthermore, Lykken (1981) advocates the application of strict psychometric principles in developing test items so that proportions of false positives and false negatives are known statistically and procedures for administering the test are uniform. Several analog studies have demonstrated that the accuracy of this test in identifying guilty subjects ranges from 70% to 100% (mean of about 88%, Lykken, 1985) and that the false-positive rate is below 5% (Lykken, 1985; Yohman, 1978). However, it has not been used widely in the field. This may reflect the reasonably stringent requirements of the test (a) to keep important details of the crime from being revealed to the public in general and to the various suspects in particular and (b) for the investigator to detail the circumstances of the crime for subsequent use in the test (see Lykken, 1981, for a description of one successful use of the GKT in a criminal investigation and Lykken, 1988, for a refutation of (a) and (b); also see Reid & Inbau, 1977, for some clever examples of how this technique can be used in the field).

Traditional polygraphy procedures, as exemplified in the CQT, and the GKT vary both in their structure and in the assumptions on which they are based. The traditional polygraphy test derives from the assumption that there is a distinct physiologic response and a set of well-defined behaviors associated with lying, whereas development of the GKT was motivated by the assumption that autonomic arousal will be induced in the guilty subject by aspects of the transgressive act and situation known only to that person. Both the GKT and traditional polygraph tests rely, however, on the use of autonomic measures (in particular, electrodermal response) to draw inferences about guilt or innocence. To our knowledge, there is no systematic research that has identified a characteristic complex of autonomic and behavioral responses indicative of fear-induced arousal secondary to lying, thereby undermining the assumption on which traditional polygraphy is based (Kleinmuntz & Szucko, 1982, 1984; Lykken, 1959, 1981, 1984; Ney, 1988; Saxe et al., 1985)[2] If recognition of critical details pertinent to the circumstances surrounding the transgressive act is the key element in detecting guilt, then psychophysiological measures that are more direct manifestations of this neurocognitive process may provide better means by which to infer guilt

[2] This criticism and its implications for policymaking are summarized by Saxe, Dougherty, and Cross (1985) as follows: "The most serious difficulty in the development of policy relevant conclusions about polygraph testing is the lack of theory to explain results of testing. Theoretically, a polygraph test operates on a simple principle: that anxiety is related to lying. Yet, it is clear that anxiety has a host of causal factors and making attributions of deception is cognitively complex (cf. Ekman, 1985). It is for this reason that many scientists are skeptical about claims for polygraph testing" (p. 364).

or innocence than can the peripheral autonomic measures used in the GKT. The analysis of changes in the properties of components of the event-related brain potential measured at the scalp suggests itself under these circumstances. Recent work indicates that these more direct measures of cognitive engagement may indeed provide a very powerful method by which to infer guilt. Before describing this work, we present a brief tutorial on cognitive psychophysiology.

Event-Related Brain Potentials

We regard the problem of detecting deception or inferring guilt[3] on the basis of a neurocognitive link between recognition of critical stimulus events and the acknowledgment of that recognition to fall naturally within the purview of cognitive psychophysiology: the area of psychophysiology devoted to studying human cognition using measures of brain electrical activity. Cognitive psychophysiologists study two classes of dependent variables: traditional behavioral measures, such as response latency (i.e., reaction time [RT]) and accuracy, and measures of brain electrical activity recorded at the scalp, commonly referred to as *event-related brain potentials* (ERPs). ERPs are recorded by affixing electrodes on the subject's scalp, typically at placement sites used throughout the world for both clinical and research purposes (10–20 or International Electrode Placement System; Jasper, 1958). These potentials represent positive or negative deflections (called *components*) in the voltage of the ongoing electroencephalographic (EEG) signal that are produced by the presentation of a stimulus to which the subject either sits passively or makes a decision leading to either an overt (e.g., button press) or covert (e.g., running mental count) response.

Components of the ERP elicited under these experimental conditions are labeled on the basis of their electrical polarity and the minimum latency at which their peak amplitude is reached (Donchin, Ritter, & McCallum, 1978). For example, N100 refers to a negative change in voltage whose peak is achieved at about 100 ms after presentation of a stimulus. Analyses are done most often on the amplitude, latency, and scalp distribution (i.e., relative amplitudes at different scalp sites) of these potentials. Because the voltages of ERPs are small in relation to those of the EEG, these signals may be obscured by the background EEG and are most evident in a signal that represents the average of a large number of repetitions of the stimulus. However, multivariate techniques such as discriminant functions and lag cross-correlations have allowed the detection of larger ERP components on single trials with a high degree of success (Horst & Donchin, 1980; Squires & Donchin, 1976). To do these single-trial analyses, however, the digitized signal must be filtered off-line to extract the signal from the noise (Farwell, Martinerie, Bashore, Rapp, & Goddard, in press; Ruchkin & Glaser, 1978).

Exogenous and Endogenous Components of the ERP

Exogenous components. The earliest components of the ERP are obligatory electrical responses to stimulation recorded typically in circumstances designed to assess the integrity of primary afferent pathways (i.e., for clinical evaluations) or to characterize factors that influence transmission rates in these systems (e.g., age and body temperature). Their production is dependent on external stimulation, and the amplitudes and latencies of these signals vary, largely as a function of the physical properties of the eliciting stimulus (e.g., an increase in the intensity of the stimulus will produce an increase in the amplitude and decrease in the latency of the component to asymptote). Hence, these components are often referred to as *exogenous* or *evoked potentials.* Moreover, the scalp site where the amplitude of the evoked potential is largest and the latency is shortest is determined by the modality stimulated (e.g., a visual stimulus evokes the most robust response over occipital—primary visual cortex—recording sites). Consequently, electrode placement is determined by the sensory modality in which the stimulus is delivered.

Endogenous components. When decisions must be made about the stimuli presented (e.g., count a high-pitched tone, refrain from counting a low-pitched tone), exogenous components are manifest within the first 100–150 ms after the stimulus. At longer latencies, a series of components referred to as *endogenous* is evident. The amplitude, latency, and scalp distribution of these components of the ERP are relatively insensitive to changes in the physical properties of the stimulus. Instead, their amplitude and latency are very sensitive to changes in the psychological demands of the task, and their scalp distribution is relatively constant across tasks and sensory modality. Thus, these components are believed to be generated by neural events associated with higher order processing of the stimulus input and response output. Hence, the label *endogenous* for these components, examples of which are the N200, P300, and N400. Because the scalp distribution of the endogenous components is not tightly coupled to the sensory modality in which stimulation occurs but may vary with the cognitive demands of a task

[3] In the debate that rages over the validity of approaches to polygraphy that depend on autonomic measures, deception is distinguished from guilt. As we have discussed, the former is presumed to be manifest in a pattern of physiological activation associated with the fear induced by explicitly lying (assessed in the CQT) and the latter with physiological arousal evoked by the recognition of critical crime-relevant information (assessed by the GKT). This distinction is essential in differentiating emotion-based (CQT) from recognition-based (GKT) autonomic approaches. It may, however, be less important for the alternative approach that is based on event-related brain potentials, which we next discuss. That, of course, is an empirical issue to be resolved. Deception, in our view, includes not only overt acts of lying like those compelled of the guilty person in the CQT but also covert attempts to conceal recognition of critical pieces of information such as those that the guilty person may try to engage in the GKT (e.g., attempts to relax to the presentation of each stimulus). Bok (1978) argues that lying includes, for example, not only explicit denial of actions for which one is culpable but also the maintenance of silence in circumstances in which an overt expression of culpability is either not required or is optional. We conceive of deception in this broader sense and, consequently, would include both the CQT and the GKT as methods for detecting deception. As is apparent, however, we view the neurocognitive–information-possession link to be theoretically the most viable approach to inferring guilt or innocence. This article is not the appropriate forum to debate the concept of deception. Accordingly, and in deference to tradition, we abide by the deception–guilt distinction in our discussion.

(e.g., be more or less lateralized), it is not unusual for a large array of electrode placements to be used in investigations of these components. Electrodes are usually placed at locations along the midline of the scalp (referred to as Fz [frontal], Cz [central], Pz [parietal], and Oz [occipital]), and at a variety of lateral recording sites (e.g., temporal sites T3, T5; central sites C3, C4; and parietal sites P3, P4). Figure 1A shows examples of ERPs with their constituent exogenous and endogenous components.

Note that the neural origins of the early evoked potentials are characterized far better than are those of the late endogenous components (Goff, Allison, & Vaughan, 1978). Indeed, it is not unfair to assert that the sources of the late potentials are unknown (Wood et al., 1984). It is fair to claim, however, that the elements of mental processing indexed by a variety of these endogenous ERPs are becoming increasingly well characterized (Donchin, Karis, Bashore, Coles, & Gratton, 1986; Picton & Stuss, 1980; Ritter, Vaughan, & Simson, 1983). In the following discussion, our attention is focused on the P300 and N400, the two components that we believe have the greatest potential for use in interrogation procedures designed to differentiate guilt from innocence.

The P300. Early investigations of the P300 determined that it is manifest in simple information-processing tasks in which the subject is required to monitor a series of stimulus events comprising two stimuli, one of which signals an overt (e.g., button press) or covert (e.g., running mental count) response and the other of which indicates that no response is required (for reviews see Donchin, Karis, Bashore, Coles, & Gratton, 1986; Duncan-Johnson & Donchin, 1982; Johnson, 1986; Pritchard, 1981). Whenever the critical event is presented, a P300 is elicited, the amplitude of which is inversely related to the probability of occurrence of that event. That is, the less frequent this event, the larger the amplitude of the P300 it elicits. Indeed, it is now well established that the amplitude of the P300 is determined by the task relevance of the stimulus and by the frequency of occurrence of the critical event, so that relevant stimuli produce P300s that are largest when the event is rarest. More precisely, the amplitude of the P300 is tied closely to the subjective probability of the target event (see Donchin & Coles, 1988, for an in-depth discussion of this relationship). The sensitivity of P300 amplitude to variations in the probability of the target stimulus is shown in Figure 1B.

The finding that low-probability events elicit the largest P300s has given rise to the name *oddball* for the basic task in which the P300 is generated. In this task, the subject is presented with one of two possible stimuli (e.g., 1000-Hz and 2000-Hz tones) on a trial, an infrequent stimulus (e.g., $p = .15$) designated as the target, and a frequent stimulus identified as the nontarget. The subject's task is to keep a running mental count of the rare target stimulus (or to press a button when it occurs) but to keep no mental count of presentations of the frequent nontarget (or to withhold the button press). A large P300 is elicited by the counted stimulus, and little or no P300 is produced by the uncounted stimulus. As the probability of occurrence of the target increases, the amplitude of the P300 decreases, and a P300 begins to emerge to the nontarget. At equiprobability, both stimuli produce a P300, but the P300 to the target is slightly larger. This is the target effect. Most of the early work on P300 used the oddball task or some variant of it. More recent work has shown that the P300 is elicited in complex information-processing tasks in which decisions must be made and action initiated about relevant classes of stimulus events (e.g., Bashore & Osman, 1987; Callaway, 1983; Coles, Gratton, Bashore, Eriksen, & Donchin, 1985; Donchin, 1981; Donchin, Karis, Bashore, Coles, & Gratton, 1986; Duncan-Johnson & Donchin, 1982; Kutas, McCarthy, & Donchin, 1977; Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981; Mulder, Gloerich, Brookhuuis, van Dellen, & Mulder, 1984; Ritter et al., 1983). Hence, the P300 is manifest not only under conditions in which the subject must distinguish the target from the nontarget stimuli on the basis of a simple physical property (e.g., high- vs. low-pitched tone; red vs. green light) but also when the subject must categorize the relevant stimulus events on the basis of an abstract property (e.g., count the names of males and refrain from counting the names of females).

It has also been shown that the scalp distribution of the P300 is reasonably consistent across stimulus modalities and information-processing tasks; in young adults, it is largest in amplitude at Pz, intermediate at Cz, and smallest at Fz. Note that the designation 300 refers to the latency of this component in tasks using simple stimuli with easy discriminations between targets and nontargets. As processing difficulty increases, the latency of the P300 can be prolonged by as much as 300 or 400 ms. Under these circumstances, identification of this component as a P300 is done on the basis of its scalp distribution and responsivity to experimental manipulations.

An interesting demonstration of the endogenous nature of the P300 is provided in the omitted-stimulus task (Ruchkin, Sutton, & Tueting, 1975). In this task, the subject is presented a series of identical stimuli (e.g., 1000-Hz tones) with an interstimulus interval of 1,000 to 1,500 ms that is interrupted infrequently (e.g., 15% of the trials) by an omission of the stimulus. The subject's task is to count occurrences of this omission. The ERP generated to the omission contains a large P300, whereas the ERP produced by the stimulus does not. Thus, the P300 can be generated in the absence of an evoking physical stimulus. Again, the importance is that the omission is categorized as relevant.

As is apparent in the preceding, the conditions necessary for invoking the P300 have been reasonably well characterized. If this brain event is to be used to infer recognition of critical information-processing events, however, those elements of mentation that are engaged when the P300 is generated must be known. A fair amount of research in the past 20 years has been devoted to achieving this end. Ritter, Simson, and Vaughan (1972) and Kutas et al. (1977) suggested that the latency of the P300 reflects the time taken to evaluate and categorize stimulus events and is relatively independent of the time required to select and execute a response. In support of this hypothesis, Kutas et al. (1977) reported that the latency of P300 increased as the difficulty in classifying a stimulus increased and that variations in response set (speed vs. speed and accuracy instructions) produced large changes in RT but little change in P300 latency. This dissociation between the effects of response set on RT and P300 latency has also been reported by Pfefferbaum, Ford, Johnson, Wenegrat, and Kopell (1983), Coles et al. (1985), and Strayer, Wickens, and Braune (1987).

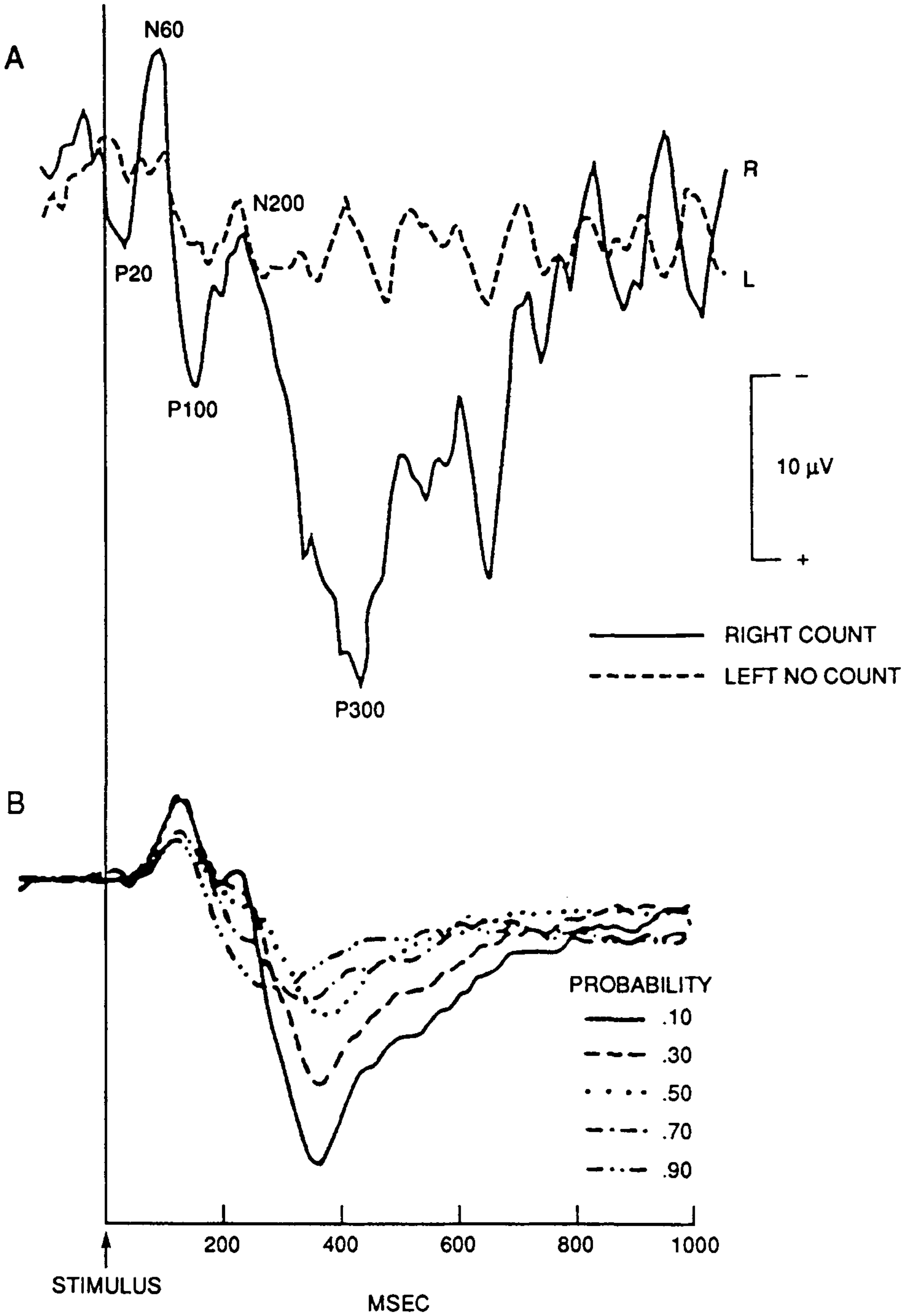

Figure 1. (A) Average event-related brain potentials (ERPs) elicited in an oddball task in which the subject was delivered a mild electric shock to the left or to the right index finger. (She was required to count stimulation of the right index finger and to refrain from counting stimulation of the left index finger. Stimulation of either finger occurred randomly but was delivered to the right index finger only on 20% of the trials. These averages were recorded in a block of 100 trials [20 right, 80 left]. The arrow indicates the onset of the stimulus. Exogenous [P20, N60, P100] and endogenous [N200, P300] components are identified in the ERP. The emergence of the P300 to the target stimulus is clearly evident in this subject. From "Chronopsychophysiology: Mental Chronometry Augmented by Physiological Time Markers" [p. 89] by M. W. van der Molen, T. R. Bashore, R. F. Halliday, and E. Callaway, 1991, in J. R. Jennings and M. G. H. Coles, *Handbook of Cognitive Psychophysiology,* New York: Wiley. Copyright 1991 by John Wiley & Sons, Ltd. Adapted by permission.) (B) The average ERPs for the effect of varying target probability on the amplitude of the P300 in the auditory modality. (Subjects were required to count one tone and to refrain from counting another tone. The recordings shown in Part A and Part B were all made at the parietal electrode site. From "A Triarchic Model of P300 Amplitude" by R. Johnson, 1986, *Psychophysiology, 23,* p. 369. Copyright 1986 by The Society for Psychophysiological Research. Adapted with permission of the author and the publisher.)

A very convincing demonstration of the relationship of P300 latency to stimulus processing was provided by McCarthy and Donchin (1981). They varied stimulus discriminability and response compatibility orthogonally and found that the stimulus manipulation influenced both P300 latency and RT but that the response manipulation altered RT and had little effect on P300 latency. The inference drawn from this study was that the timing of P300 is influenced strongly by stimulus processing and only weakly, if at all, by response processing. This basic finding has been replicated and extended in work by Magliero et al. (1984). Other demonstrations of the dependence of P300 latency on the rate of stimulus processing have been provided by Callaway (1983), Ford, Roth, Mohs, Hopkins, and Kopell (1979), and Mulder et al. (1984).

In summary, generation of the P300 appears to depend on the categorization of relevant stimulus events, to vary in timing as a function of the difficulty of this categorization, and to vary in amplitude as a function of the probability of occurrence of the critical stimulus.

The N400. The N400 component of the ERP was first identified by Kutas and Hillyard (1980c). They presented subjects with a series of seven-word sentences, one fourth of which ended with an unexpected word (e.g., "He shaved off his mustache and eyebrows") and the rest with an expected word (e.g., "I take coffee with cream and sugar"). Sentences that ended with an expected word produced a P300, whereas those that were completed by an unexpected word elicited the N400. Like the P300, the N400 had a predominant centroparietal maximum. Further research by Kutas and others revealed that the N400, like the P300, is extremely robust and is readily invoked by semantic anomalies in both the visual and the auditory modalities (Holcomb, 1988; Kutas & Hillyard, 1980a, 1980b, 1984a, 1984b; McCallum, Farmer, & Pocock, 1984; see review in Kutas & Van Petten, 1988); studies in the visual modality demonstrated that it can be elicited in French (Besson, Macar, & Pynte, 1984), Spanish (Kutas, 1985), and American Sign Language (Kutas, Neville, & Holcomb, 1987; Neville, 1985). Indeed, the amplitude of N400 seems to vary systematically with the degree of semantic anomaly; the larger the disparity, the bigger the N400. The anomalous word does not have to appear at the end of the sentence, however. It can be produced by semantically anomalous words that occur in the middle, as well as the end, of a sentence (Kutas & Hillyard, 1983). Unlike the P300, whose amplitude is quite sensitive to variations in the probability of the critical stimulus events, the amplitude of the N400 does not appear to vary with probability. Moreover, it seems to be quite specific to deviations within the linguistic domain. Deviations in melodies, geometric shapes, and notes of the musical scales (Besson & Macar, 1987), presentation of novel, unexpected line drawings (Kutas & Hillyard, 1980c), and presentation of physical (Kutas & Hillyard, 1980c) or grammatical (Kutas & Hillyard, 1983) anomalies do not produce an N400. They produce, instead, a P300.

Other research has revealed that the N400 can also be elicited by single words when they are primed by a preceding word and that prime is violated (Kutas & Hillyard, 1989). Although an N400 can be elicited under these circumstances, its amplitude is larger to words that end a sentence. This may reflect its hypothetical relationship to priming. Single words that end sentences anomalously may produce larger N400s than single words following a single-word prime because the preceding-sentence context may have primed an expected final word to a greater degree than does the single-word prime and a violated expectancy is proportionately greater.

As suggested in the preceding, the P300 and N400 are manifest when expectations are violated (i.e., to deviations). The conditions that elicit these two components differ, however. The N400 appears to be specific to language or some aspect of language processing, whereas the P300 is a more general response to infrequent or task-relevant stimulus events, both in and out of the language domain. Work by Kutas and Hillyard (1980c) suggests that the two components are independent of one another but can be manifest simultaneously if both semantic and physical deviations are present in the anomalous final word of a sentence.

Of central interest to this review are the observations by Fischler and colleagues that an N400 can be elicited by incongruities in semantic, episodic, and personal knowledge. Fischler, Bloom, Childers, Roucos, and Perry (1983) assessed the extent to which N400 was manifest in a sentence-verification task. Subjects responded to a simple set of semantic propositions that were either true affirmatives (e.g., "A robin is a bird"), false affirmatives (e.g., "A robin is a vehicle"), or false negatives (e.g., "A robin is not a bird") with button presses indicating whether the statements were true or false. Fischler et al. (1983) found that an N400 was produced to the final word of false-affirmative or true-negative sentences. The suggestion from this work is that production of the N400 occurs in response to unrelated words in a sentence, irrespective of the propositional content (affirmative or negative) of the statement.

In two studies with direct implications for the development of ERP-based procedures for detecting the possession of guilty knowledge (see discussion in the next section), Fischler's group demonstrated that an N400 is generated by false statements. Fischler, Childers, Achariyapaopan, and Perry (1985) had subjects memorize a set of statements whose veracity was defined by the experimenters. Examples of true statements would be "Diane is a chemist" and "Matthew is a lawyer"; examples of false statements would be their recombinations, "Diane is a lawyer" and "Matthew is a chemist." Subjects were drilled on these statements until they could discriminate the true from the false with near-perfect accuracy. In a test session the next day, ERPs were recorded as the subjects (in separate blocks) either indicated that the statements were true or false by pressing the appropriate response button or simply read the statements without making any overt response. False statements in both cases produced an N400. In a related study, Fischler, Bloom, Childers, Arroyo, and Perry (1984) assessed ERP responses to self-referential statements that were either true or false. These statements were either strongly (e.g., "My name is . . .") or weakly (e.g., "My favorite author is . . .") self-referential. Statements were made false by substituting a semantically correct alternative. Fischler et al. (1984) found an N400 to both strong and weak false statements, the amplitude of which did not vary as a function of the strength of the self-reference. These results suggest that the N400 is sensitive not only to semantic anomalies but also to associations or relationships within semantic and episodic memory.

In summary, current research suggests that the N400 is produced by deviations within a sequence of stimuli only if the context and deviation are linguistic. This component also appears to be produced in response to stimulus events from both semantic and episodic memory that are recognized as false by the subject.

ERPs as Tools for Inferring the Possession of Guilty Knowledge

We are aware of only a small number of published reports in which alterations in the properties of ERP components have been analyzed to infer the possession of critical pieces of information (Farwell & Donchin, 1986, 1988a, 1988b, 1991; Rosenfeld, Cantwell, et al., 1988; Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987). In each of these studies, the P300 was the component of interest, and variations in its amplitude were used to infer the possession of guilty knowledge. Variants of the GKT were developed for use in each of these studies that took advantage of the P300-eliciting properties of the oddball task. The reasoning that motivated this work was as follows: If P300 represents a manifestation at the scalp of neural events associated with the recognition and discrimination of critical stimulus events and is influenced little by response output processes, then it is conceivable that this component will emerge under circumstances in which a person recognizes a stimulus but attempts to conceal that recognition. In other words, it should be manifest in conditions similar to those in the GKT.

Rosenfeld and associates (Rosenfeld, Cantwell, et al., 1988; Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeiri, 1987) had subjects select an item (e.g., a wallet) from a box of nine items and then showed them a sequence of names of items on a computer screen that included the selected item and the names of items that were not in the box. In the first experiment, subjects were instructed to count the selected item, and in the second experiment they were instructed to try and beat the system by counting another item. These are both variants of the classic oddball task that require a covert response by the subject: mental counting. In both cases, a large P300 was elicited by the selected item but not by the ignored novel items. In the second study, both the counted item and the selected item produced P300s that were larger than those produced by the novel items. The results of the first experiment simply demonstrate that a counted infrequent target stimulus produces a large P300, whereas those of the second experiment suggest the possibility that information the subject is trying to conceal may be revealed in the P300. Unfortunately, the instruction to select another item to count and to ignore the selected (infrequent) item may actually promote it as a target, not unlike asking a person not to think about the rare white monkey.[4]

That this may not be the case, however, is suggested in the work of Farwell and Donchin (1991). Their subjects completed an interactive computer briefing in which they learned the details of a mock espionage scenario. After this briefing, subjects then carried out the scenario. They went to a designated location and passed sensitive information to the person they met. Each scenario had five specific pieces of information associated with it. The next day, subjects were tested for their knowledge of the details of two scenarios: one in which they had participated and were aware of the details and another in which they had not participated and were unaware of the details. Subjects were thus informed (i.e., guilty) as well as uninformed (i.e., innocent) about the details of a scenario. They were presented a sequence of stimuli, each of which was composed of 2 one-syllable words, defined by the experimenters as targets, to which subjects were to press one response button, or as irrelevant nontargets, to which they were to press another response button. The targets and nontargets occurred randomly on 17% and 66% of the trials, respectively. A third stimulus category existed as well, labeled by Farwell and Donchin as *probes.* These stimuli referred to critical details of a scenario known only by the informed participant; they appeared randomly on 17% of the trials, as did the targets. Note, however, that unlike the Rosenfeld studies, the probes were not distinguished in the instructions given to the subjects. Rather, the subjects were simply informed of the target and nontarget distinction. Hence, if the probe items elicited a large P300, it could be argued that these items were implicitly categorized as salient by the subject. That is, the possession of guilty knowledge would be expressed in the emergence of P300s to the pertinent probe-stimulus events.

Farwell and Donchin (1991) predicted that the standard pattern of results would be evident in all subjects under both conditions (informed or uninformed) to the target and nontarget stimuli: a large, distinct P300 to the targets and little or no P300 to the nontargets. The predicted response to the probes would vary, however, as a function of the subject's knowledge of the scenario details. When aware, the subject was expected to treat probes in a manner akin to targets, thereby generating P300s of comparable magnitude to the two classes of stimuli. When unaware, the subject was expected to treat the probes as nontargets and to generate little or no P300 to either. Very strong support was provided for these predictions. As expected, the P300s elicited by the target stimuli were large and distinct and easily differentiable from the positivity elicited by the nontarget stimuli. However, the essential comparisons were between the P300s elicited by the probes and the other two classes of stimuli: P300s to the probes would resemble those to the target when the subject was informed of the details of the scenario but would resemble the positivity associated with the nontargets when uninformed of the details. If this pattern prevailed, guilt or innocence could be inferred. This is precisely what Farwell and Donchin (1991) found, using a bootstrapping technique that established correlation distributions between the P300s elicited by the target, nontarget, and probe stimuli at the midline parietal scalp site for each subject on a matched number of single trials. Of the 40 determinations thus derived (20 subjects, informed and uninformed), only 12.5% were indeterminate. The balance were all correctly classified as guilty or innocent, yielding validities of 100% (excluding inconclusives) and 87.5% (including inconclusives). The ERPs recorded from these subjects in each condition are shown in Figures 2 and 3.

In a second experiment, Farwell and Donchin (1991) as-

[4] The rare white monkey example is used to make the point that when a person is asked not to think about (i.e., not to attend to) something, he or she will invariably think about (i.e., attend to) it. This example did not originate with us, but we cannot recall its source.

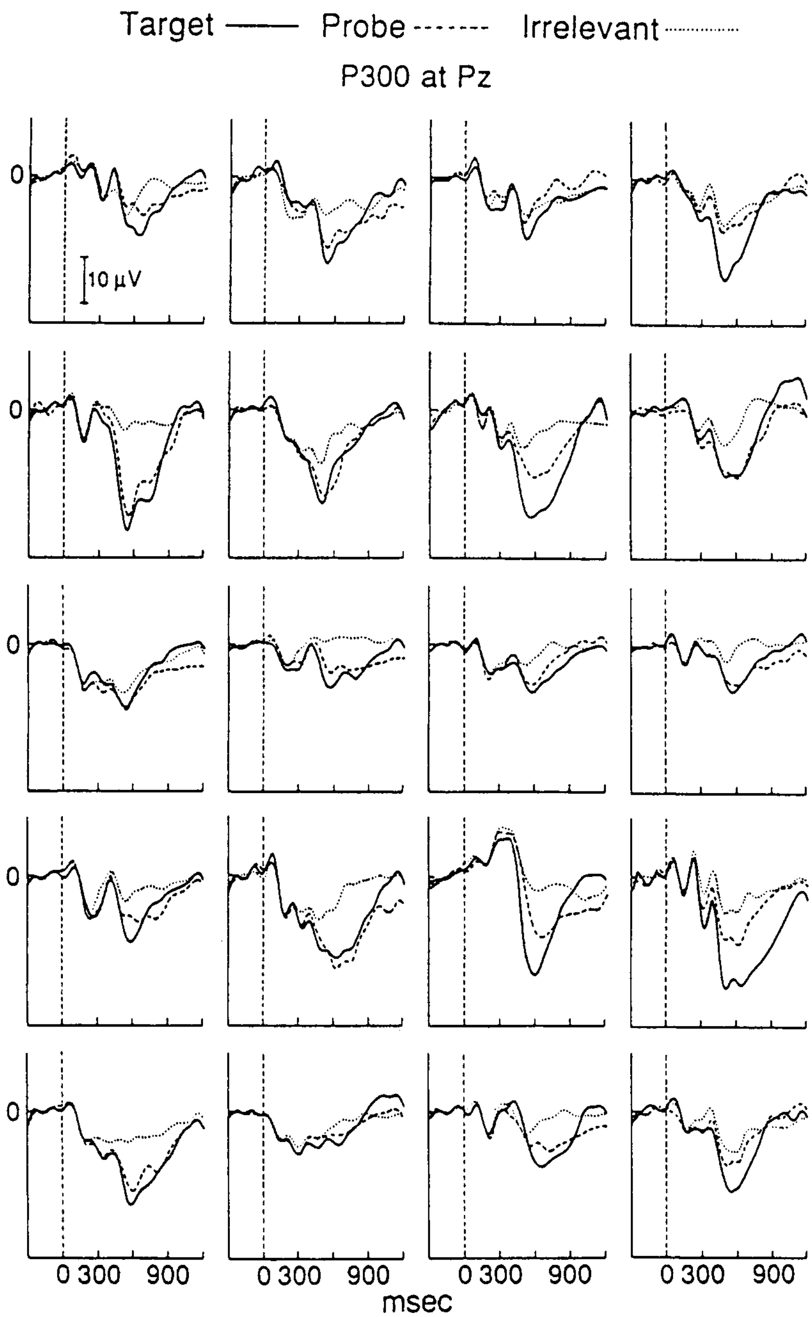

Figure 2. The average event-related brain potentials recorded at the parietal electrode site (Pz) by Farwell and Donchin (1991) from 20 subjects when they possessed guilty knowledge. (The solid, dashed, and dotted lines represent brain activity invoked by the target, irrelevant, and probe stimuli, respectively. P300 = an endogenous component. From "The Truth Will Out: Interrogative Polygraphy ['Lie Detection'] With Event-Related Brain Potentials" by L. A. Farwell and E. Donchin, 1991, *Psychophysiology, 28,* p. 536. Copyright 1991 by The Society for Psychophysiological Research. Reprinted with permission of the author and the publisher.)

sessed the extent to which this pattern of results could be replicated when the probe stimuli were sampled from information pertinent to a crime that the subject had actually committed a considerable time before testing. Four undergraduates who had *admitted to engaging* in minor criminal or socially undesirable activities (e.g., arrested for underage drinking) served as subjects. They were tested in a task that only varied in two ways from the task used in the first experiment: (a) The probe stimuli were relevant to the identified transgression, and (b) rather than make designated button press responses to each stimulus

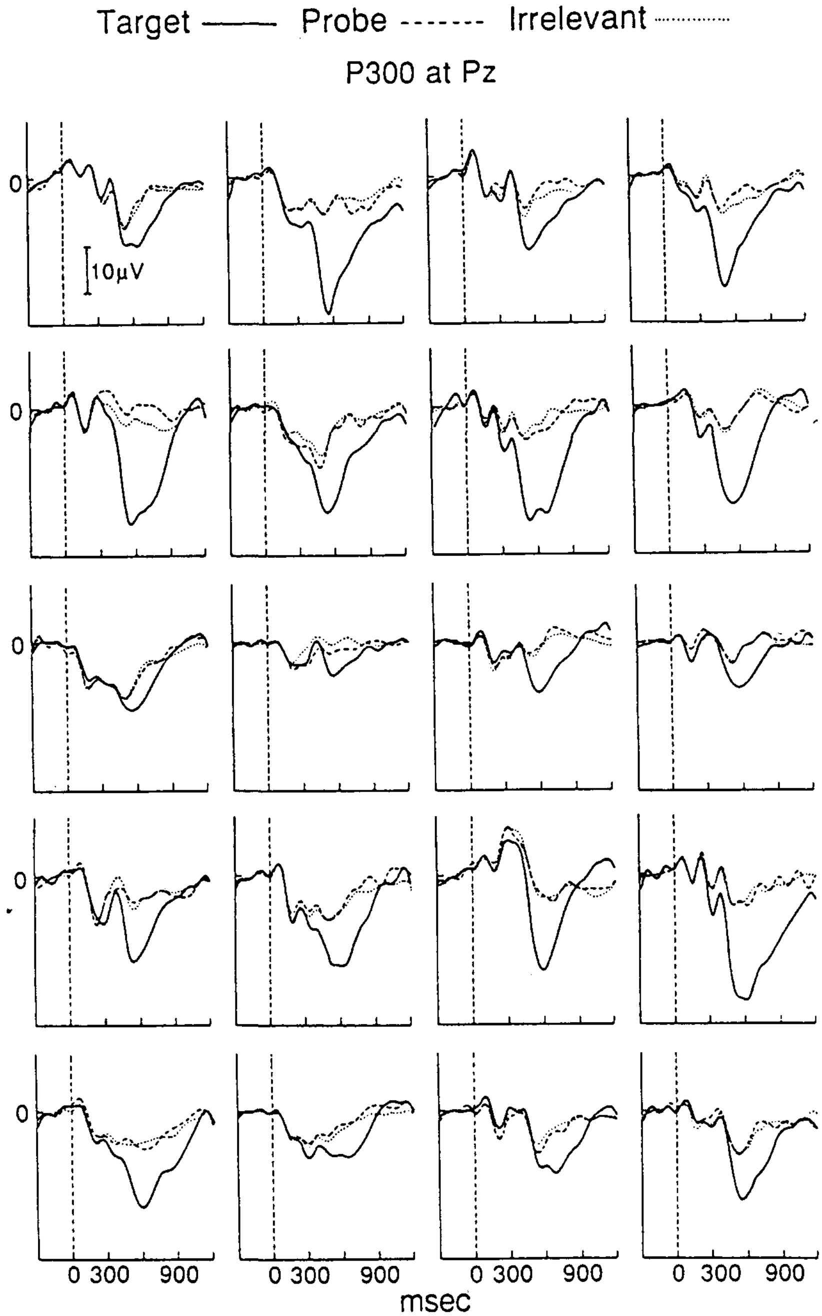

Figure 3. The average event-related brain potentials recorded at the parietal electrode site (Pz) by Farwell and Donchin (1991) from 20 subjects when they did not possess guilty knowledge. (The solid, dashed, and dotted lines represent brain activity invoked by the target, irrelevant, and probe stimuli, respectively. P300 = an endogenous component. From "The Truth Will Out: Interrogative Polygraphy ['Lie Detection'] With Event-Related Brain Potentials" by L. A. Farwell and E. Donchin, 1991, *Psychophysiology, 28,* p. 537. Copyright 1991 by The Society for Psychophysiological Research. Reprinted with permission of the author and the publisher.)

presentation, these subjects simply kept a running mental count of the occurrences of the target stimuli and refrained from counting the other stimuli (i.e., the nontarget and probe stimuli). The pattern of results reported in the first experiment was replicated exactly in this experiment. Thus, a large P300 was elicited by the uncounted probe stimulus that resembled the P300 invoked by the counted target stimuli when the probes were taken from an actual transgression by the subject (i.e., when the subject was guilty) but not when they were taken from an unrelated incident (i.e., when the subject was innocent).

The two experiments reported by Farwell and Donchin (1991) provide very strong demonstrations of the potential of P300 in inferring the possession of guilty knowledge under conditions in which the subject was not instructed to conceal that knowledge. In the traditional P300 task, the experimenter arbitrarily defines the salient stimulus catagories, as did Farwell and Donchin (1991) for the targets and nontargets, and the subject is obliged to sort the stimuli accordingly. Farwell and Donchin's work has revealed, however, that a person's categorization of stimulus events, at least into relevant and irrelevant, may be bound in an obligatory way to prior knowledge the person has of those events. Thus, the inference is supported that subjects distinguish between various stimulus categories, without being instructed to do so, when information available to them permits such distinctions to be made. This neurocognitive link may be engaged automatically (i.e., out of the person's control), irrespective of either the person's covert (e.g., ignoring the critical events by not counting them in an oddball task) or overt (e.g., pressing a button to indicate that the relevant event is irrelevant) expression to the contrary. If so, this provides a very powerful bridge to detecting the possession of critical information within the context of a GKT. It suggests that the neurocognitive mechanisms mediating the categorization of critical stimulus events cannot be disengaged by the intention to recategorize these events if this recategorization is disingenuous.

Farwell and Donchin (1991) argued that the context in which an interrogation occurs is crucial and implied that the most powerful techniques may require embedding the interrogation, as they did, within the context of another, seemingly unconnected, task. This may be the case. However, the same pattern of results they obtained by concealing the intent of a simple oddball task may be obtained with more standard methods of interrogation of subjects who are very clearly aware of the need to be deceptive if guilty. The larger P300 literature suggests this possibility. First, the P300 apparently originates from neural sources that are engaged automatically by the recognition of critical stimulus events (see review in Johnson, 1988). Second, this engagement, as revealed elegantly in the Farwell and Donchin (1991) study, may be independent of the overt expression of such recognition (see reviews in Donchin, Karis, Bashore, Coles, & Gratton, 1986; Duncan-Johnson & Donchin, 1982). And, third, the amplitude of P300 appears to vary as a function of the relevance of the stimulus to the subject (Johnson, 1988). These properties of the P300 suggest that it may be sufficiently robust to obviate the need to conceal the interrogation within the context of some broader task. Thus, the P300 may be elicited in situations in which people are informed of the object of the interrogation and they deny recognition of critical stimulus events known to them. Indeed, the relevance of stimulus events to them may be revealed in gradations of P300 amplitude. If so, it will not matter if the context of the interrogation is concealed from or known by the subject or if the subject is told to be deceptive or not. What will matter most is that the subject's brain automatically expresses the recognition of critical stimulus information, irrespective of what the subject is willing to reveal.

That this assumption has merit is suggested in a recent study by Allen, Iacono, and Danielson (1992). These investigators assessed the extent to which characteristics of the ERP, used either alone or in conjunction with behavioral measures, revealed a person's memory for specific information that he or she did not admit to remembering. Their approach involved the use of Bayesian posterior probabilities to classify ERPs recorded in a conventional oddball task. Subjects in this study were first required to learn a list of words from a category they selected themselves (from among seven choices). These words then served as the target stimuli in an oddball task that included four other categories of words (not from among the previous choices), each of which occurred with a random, but equal, probability. After a break of 30 min, the subjects were required to learn a new list and then complete another oddball task. They were not informed of this second task until it was begun. The words in the new list served as the target stimuli, and the previously learned list of words, along with five other new category lists, served as the nontargets. Again, each category occurred with random, but equal, probabilities. In the first experiment (of three), subjects were instructed to ignore the previously learned words; in the second and third experiments, they were instructed to respond honestly to the most recently learned words (i.e., to treat them as targets) but to lie about having learned the first set of words (i.e., to treat them as nontargets). They were encouraged to attempt to prevent their brainwaves from letting the investigators know that they were lying. In the third experiment, a $5 incentive was offered for fooling the system. Subjects responded to the stimulus with a button press indicating whether it was a target (e.g., left button press) or nontarget (e.g., right button press). In each experiment, near-perfect (94% across the three experiments) identification of learned material as learned (i.e., recognized) and rare (4% across the three experiments) misclassification of unlearned material as learned were achieved on the basis of ERP morphology differences. The instructional set did not alter the classification accuracy of the Bayesian procedure. That is, encouragement of the subject to attempt to deceive the system had no apparent effect on the classification process. Thus, differences in ERP morphology to critical versus noncritical pieces of stimulus information may reveal themselves even in circumstances in which the intent of the ERP-based interrogation is revealed.

Research Issues in Polygraphy Using ERPs

Although suggestive, the ERP studies we reviewed represent the earliest efforts in what promises to be a lengthy research endeavor. This research effort will have to grapple with all of the problems that have plagued traditional polygraphy as well as those that are unique to ERP research. As has been demonstrated in this and numerous other reviews, there is ample reason to be skeptical of the validity of both the behavioral and physiological observations used in traditional interrogative polygraphy. Despite this, opponents and proponents alike of these procedures persist in using autonomic measures (in particular, the electrodermal response) as the sole physiological indexes of deception. Distinguished proponents such as David Raskin and his colleagues eschew behavioral observations and seek to develop more refined numerical scoring of the polygraph record. They do not question the validity of the CQT or the assumptions on which it is based. Distinguished opponents such as David Lykken challenge both the technique and its assumptions but continue to use autonomic measures of arousal to infer guilt.

Indeed, with the exception of the ERP studies we reviewed, the research done in the field uses peripheral measures. It is conceivable, however, that refinement of the questioning technique, so that it meets strict psychometric criteria, and numerical standardization of the measurement procedures cannot overcome the problems inherent in using peripheral physiological measures. These problems are summarized eloquently by Waid and Orne (1981), even though the point of their discussion is to describe variables that have been neglected but probably should be controlled in traditional polygraphy. It is clear from their review that the validity of this methodology depends on more than adherence to sound psychometric principles and standardized scoring. It depends as well, in their view, on the control of a broad range of cognitive, social, and personality variables. We agree but think that any review of traditional polygraphy should be taken a step further— to suggest the use of alternative physiologic indexes of guilt or deception that are more direct measures of the processes of interest, are more stable, and are less influenced by nonphysiologic (i.e., situational) variables. As we attempted to demonstrate, measures of brain electrical activity are quite viable alternatives. In this section, we describe variables that can influence the outcome of traditional polygraphic diagnoses and discuss their effect on brain electrical activity. We also discuss issues that are unique to the use of ERPs in interrogative procedures. The entire thrust of this discussion is to identify variables that, in our view, must be studied systematically so as to develop procedures that minimize the proportion of false positives and maximize the number of true negatives.

Analog Versus Field Studies

Studies of the effectiveness of autonomic measures to differentiate guilt from innocence fall into two domains, field or analog (see review in the OTA, 1983, report). The former includes studies of the results of polygraphic examinations given by professional polygraphers in actual criminal investigations using accepted polygraph techniques (also referred to as *real* studies), whereas the latter comprises studies in which polygraphic techniques are evaluated under controlled experimental conditions with volunteer subjects who typically enact appropriate parts in mock crimes and then complete a polygraph examination about their involvement in the staged crime (hence, these studies are also referred to as *laboratory* studies). In field studies, the outcome of the polygraphic examination is evaluated against some post hoc determination of the subject's guilt or innocence (e.g., a judicial decision or a confession), referred to as the *ground truth.* Many proponents of traditional polygraph techniques assert that only under field conditions can a valid assessment of these procedures be conducted. Critics of this type of study argue, however, that its outcome can be confounded by the fact that the polygraph examination may vary idiosyncratically from one examiner to another, that the polygrapher may have had prior exposure to information that was critical to arriving at the judicial decision of guilt or innocence, and that in actual criminal investigations suspects may be falsely convicted or may confess to a crime that they did not commit, thereby making it extremely difficult to determine ground truth. In contrast, although an artificial situation, the analog study permits the investigator to establish ground truth and to systematically control the conditions of the polygraph examination. Critics of laboratory studies counter that the circumstances under which these studies are conducted are fundamentally different from those in the field, rendering them of questionable validity. For example, the training and competence of the examiners may differ, the subject populations may differ, and most important, the consequences of the outcome of the polygraph examination may certainly differ substantially. Indeed, the OTA report indicated that field studies were about 20% to 30% more accurate in identifying the guilty than were analog studies.

The ERP studies completed to date are all of the analog variety. Hence, those critics of this type of study in traditional polygraphy can direct their criticisms to these studies as well. Note, however, that a reanalysis of the studies assembled in the OTA report by McCauley and Forman (1988) challenged its conclusions about the relative accuracy of the two types of studies. Their analysis revealed that the difference in outcome between them could be attributed to the higher proportion of inconclusives reported in the analog studies. Indeed, they were encouraged to assert that

> average accuracy in field and analog studies could hardly be more similar once inconclusives are set aside, and we believe these results put the burden of proof on anyone who would claim superiority of detection in field polygraph testing. Results aside, it cannot even be maintained that field studies are methodologically superior because both field and analog studies have obvious methodological problems. (McCauley & Forman, 1988, pp. 80–81)

We agree with this assertion and would argue that the evolutionary development of traditional polygraph procedures has been backwards: Field development has preceded systematic laboratory evaluation of the procedures. This may reflect that the traditional procedures were developed primarily by lawyers and law enforcement officials, not by scientists. Remarkably, systematic scientific investigation of these procedures dates only to the work of Lykken and Raskin. Indeed, in his 1981 book, Lykken revealed that when he began reviewing the literature on polygraphy for his 1959 article, he was astonished to discover that essentially none existed, despite the wide use and acceptance of the procedures. Studies of the viability of ERP-based procedures in the interrogation process have begun, as they should in our view, with tightly controlled analog experiments. Under these conditions, the potential of ERPs as tools to distinguish guilt from innocence can be established, and if established, the techniques can be refined in a systematic, stepwise fashion. These procedures can then be taken into the field for evaluation. (Note that the second experiment reported by Farwell & Donchin, 1991, represents a small step in this direction.) In the next sections, we discuss some of those issues that need to be considered as the research proceeds.

Individual Differences in Physiologic Activity and Reactivity

Among the most recalcitrant problems in traditional polygraphy is the tremendous variability evident across subjects in autonomic activity and reactivity. This variability can influence the outcome of traditional polygraphic examinations. For example, subjects differ widely in the stability of their baseline electrodermal, respiratory, and blood pressure activity. People in whom this activity is labile have larger responses to relevant questions than to control questions in a CQT than do subjects with stabile responses, whether they are guilty or innocent (Waid & Orne, 1981). Hence, the likelihood of making a false-positive diagnosis on an autonomically labile subject is increased substantially. In addition, differences in autonomic reactivity among people undergoing a polygraph examination may have a similar effect on the detection of deception. Innocent subjects who are highly reactive may be more likely to be misdiagnosed as guilty than subjects who are less reactive (Horneman & O'Gorman, 1987).

The marked variability of autonomic measures and the implications of this variability for the reliability of the polygraph process leads to an important question: Is there similar variability in ERPs? Like autonomic lability and reactivity, properties of an ERP component, such as amplitude, can vary quite widely between subjects. For example, although the mean amplitude of P300 in young adults is in the range of 15–20 μV, there are subjects in whom the signal is as small as 3 or 4 μV. A reduction in P300 amplitude is particularly evident with increasing age (Bashore, 1990). In subjects with small P300s, it is difficult not only to identify the P300 but also to differentiate brain responses to targets and nontargets on the basis of amplitude differences in this component. This variation is not often reported in the ERP literature, however. Indeed, in the ERPs of some older (60 years old and over) subjects there is no clearly distinguishable P300 to the target stimulus (Marshall, Bashore, Miller, Coles, & Donchin, 1983). Marshall et al. observed this phenomenon in simple decision-making tasks (oddball, two-choice reaction time) with uncomplicated, easily discriminable stimulus events (high- or low-pitched tones, male or female names).

The problem of distinguishing the P300s elicited by different stimulus events becomes more severe in all subjects as stimulus- and response-processing demands are increased. P300 amplitude tends to reduce as processing demands are increased (e.g., see Magliero et al., 1984). It is not unusual for subjects who produce large, distinct P300s in simple decision-making tasks to produce significantly reduced P300s in more complex decision-making tasks. Also, as it becomes more difficult for the subject to discriminate target from nontarget events, differences in the amplitudes of the P300s invoked by the two classes of stimuli diminish. Thus, the problem of distinguishing the P300s elicited by different classes of stimulus events in complex information-processing tasks may prove to be an intractable one to solve for subjects who generate small P300s when processing demands are few. Of course, in traditional experimental designs that test large numbers of subjects, these few subjects may not influence the statistical pattern of results and, consequently, can be dismissed as aberrant. However, when the goal is to distinguish the responses of an individual subject to critical stimulus events, as it is in the interrogation process, these nonconforming subjects cannot be ignored.

Historically, it has been assumed that differences in the ERPs elicited by the various stimulus events will emerge in the average of a large number of stimulus presentations. Hence, until recently, analyses were typically done using values taken from the average ERP. This procedure assumes, however, that the distribution of the single-trial responses is symmetric and that the variance around the mean is stochastic: assumptions that may not be true (see critical comments in Meyer, Osman, Irwin, & Yantis, 1988). Investigators in the field have recognized this problem for a number of years and are developing procedures for extracting amplitude and latency values from single-trial data (see discussion in Coles, Gratton, Kramer, & Miller, 1986). It seems possible that the effectiveness of countermeasures, like those discussed below, will increase with each presentation of a critical stimulus. If this is so, then an ERP-based interrogative system will have to base judgments on the analysis of single-trial data. The use of single-trial data becomes especially difficult in this application because the factors listed above that affect the variability of average ERP signals exert their effects with undiminished force on single-trial signals. That is, variability that might average to zero in the averaged signal are maximally distorting to single-trial waveforms. It seems, therefore, that two factors will prove to be crucial to the success of ERP-based systems for detecting deception or inferring guilt: (a) The questions used in the interrogation must be carefully structured so that they elicit the most distinct ERP componentry, and (b) the analytic tools that are used to evaluate these responses must be able to detect subtle features in noisy single-trial data.

The first issue can probably be resolved by painstaking, systematic empirical investigation. Resolution of the second issue, single-trial analysis, is less obvious. The methods used thus far for analyzing single-trial data have met with moderate success when applied to ERPs obtained from healthy young adults in a variety of experimental paradigms. However, they have not been perfect (see the following discussion). That is, they have not been perfect when applied to the analysis of ERPs obtained under rigorously controlled experimental conditions from highly functioning, cooperative undergraduate subjects. Their success in analyzing complex, highly variable signals obtained in field conditions from potentially uncooperative, middle-aged individuals is hardly an established certainty.

Reliability and Validity

The issue of individual variation in component amplitude has important implications for the reliability and validity of ERP components, such as P300, as indexes of the processes of interest when inferring the possession of guilty knowledge. These two factors depend ultimately on the power and sensitivity of the analytic procedures used to characterize the ERP. There is a reasonably high degree of reliability in the scoring of traditional polygraph records among examiners who are trained in the same scoring technique, be it global or numerical scoring. At the same time, however, there is a large proportion of false positives. Hence, the validity of these procedures is open to very serious challenge. This disparity derives to a substantial degree from the subjective nature of traditional polygraphic scoring and the absence of a sound theoretical basis for selection of the physiologic measures of deception. Similar deficiencies need not characterize the use of ERPs in interrogative procedures.

The concept of reliability in ERP research has at least three elements. The first relates to the consistency with which the designated ERP component can be elicited from a subject by the critical test items within a test block and across test blocks. The test procedures must be sufficiently powerful to produce consistent ERP responses to the critical items both within and between test blocks. As we reviewed earlier, there is a large literature that demonstrates very convincingly that the P300 is always elicited under conditions in which the subject is called on to make categorical distinctions about stimulus events, and there is the beginnings of a literature in the identification of guilty knowledge that suggests that critical stimulus events may be categorized automatically by a subject, independent of his or her intentions. Thus, it is reasonable to assert that the P300 is produced reliably under the appropriate experimental conditions. Although in an earlier stage of its research history, the N400 appears to be a component that can be produced reliably as well. There are few studies, however, that assess the consistency of ERP components within subjects across time within a task. It has been demonstrated that the amplitude of the P300 recorded in auditory and visual oddball tasks is highly reliable when measured at two different points in time (Marshall et al., 1983). This work was done on the averaged ERP. It has been extended recently in work by Fabiani, Gratton, Karis, and Donchin (1987) in which the reliabilities of a host of measures (e.g., area, peak, cross-correlation, principal components) of P300 amplitude and latency taken from both single-trial and average ERP signals were assessed, both within and between test sessions. In general, the reliabilities exceeded .60, but did range across measures from .34 to .92. Note that more reliable estimates were derived from single trials than from averages of the ERP. This study demonstrates the importance of the measurement technique selected to derive the desired estimate of P300 amplitude or latency and suggests as well as that single-trial data may be quite useful in inferring the possession of guilty knowledge.

All of the analytic procedures depend, however, on the quality of the signal to which they are applied. ERPs must be filtered as they are being recorded (i.e., on-line) and when they are being prepared for analysis (i.e., off-line). There are accepted standards in the field for the on-line analog filtering of these signals (e.g., Coles et al., 1986). In contrast, there is no standard in the field for off-line digital filtering of the ERP and relatively little active work on identifying the optimal digital filter. The choice is nontrivial, particularly when single-trial analyses are done on small, reasonably indistinct, components and when subtle differences in ERP properties are sought. The ideal filter preserves the integrity of the signal while minimizing the contribution of noise. Thus, the second source of reliability must be found in the filtering techniques applied to the signal (for discussion of filter selection, see Farwell et al., in press).

The third influence on reliability is found in the analytic procedures selected to identify and differentiate the ERPs or specific ERP component elicited by the various test items. The procedures must be able to identify ERP responses to critical events on a consistent basis, irrespective of the absolute amplitude of the signal and the magnitude of the difference between the ERP elicited by the target and nontarget. There are a number of conventional analytic procedures of differing reliability that have been used (e.g., stepwise discriminant analysis, principal-components analysis; Donchin & Heffley, 1978; Donchin & Herning, 1975; Horst & Donchin, 1980; Squires & Donchin, 1976; see review in Coles et al., 1986) and may prove to be of value. The bootstrapping technique used by Farwell and Donchin (1991; for details, see Farwell, 1991) appears to be very promising as well. It may be the case, however, that only nontraditional analytic procedures, such as dynamical analysis (Rapp, Bashore, Martinerie, Albano, & Mees, 1989), have the necessary power to make the essential distinctions.

A considerable amount of research must be done before the validity of ERP measures in the detection of deception or the identification of guilty knowledge can be established or refuted. In our view, however, the validity of P300 amplitude as an index of cognitive processes engaged by the categorization of stimulus events into those that are relevant or irrelevant to the performance of a particular information-processing task is beyond dispute, and a strong case is emerging for N400 amplitude as an index of semantic anomaly and, perhaps, as another index of the possession of critical information. There are, however, only interesting suggestions from the work of Farwell and Donchin and Rosenfeld et al. that the stimulus processing manifested by the P300 may also provide a valid measure of guilty knowledge. Consequently, a fundamental task is rigorous demonstration of the validity of these measures in well-controlled laboratory studies that generalize to the field. In accord with Lykken and other opponents of the traditional procedures, our view is that validity is achieved only when the system of detection has a high proportion of true positives (approximating 100%) and a low proportion of false positives (approximating 0%). It is too early to determine if the use of ERPs in the interrogation process can satisfy these standards, although the work of Farwell and Donchin (1991) is promising in that it produced no false positives.

Other Variables That May Influence the Outcome of a Polygraph Examination

In addition to the problem of individual differences in autonomic lability and reactivity, a variety of cognitive, social, and personality processes can affect the relative autonomic arousal of subjects to relevant versus control (or irrelevant) questions. Some of these variables may affect components of the ERP as well.

Socialization and Other Personality Variables

Although some evidence suggests that level of socialization may be related to electrodermal reactivity (Waid, Orne, & Wilson, 1979; see review in OTA report, 1983), the bulk of the literature points to no such relationship (Balloun & Holmes, 1979; Honts, Raskin, & Kircher, 1985; Raskin, Barland, & Podlesny, 1978; Raskin & Hare, 1978). Indeed, the work of Raskin and Hare suggests that deception by psychopaths is identified in the CQT as readily as by nonpsychopaths. Likewise, there have been reports that suggest either the presence (Giesen & Rollison, 1980) or absence (Iacono, Boisvenu, & Fleming, 1984) of variation in autonomic reactivity among subjects with high and low levels of anxiety. Similarly, other personality variables have been shown to have some (Gudjonsson, 1982) or no (Iacono et al., 1984) effect on detectability in traditional polygraphy. Thus, there appears to be no strong reason to assume that a subject's personality has much impact on the examination process. Only a few studies have assessed variations in ERPs that are associated with psychopathy (e.g., Jutai, Hare, & Connolly, 1987; Raine & Venables, 1987, 1988). These studies suggest that psychopaths, at least in short duration tasks, may be better able to attend selectively to critical stimuli than do nonpsychopaths. In support of this view, Raine and Venables (1988) have reported that P300 amplitude is larger in psychopaths than in nonpsychopaths, and differences in the amplitudes of components elicited by targets and nontargets are also larger among psychopaths. The suggestion is that the possession of guilty knowledge may be identified with ERP measures more easily in psychopaths than in nonpsychopaths.

Subject–Examiner Interaction and the Nature of the Crime

The autonomic profile of a subject may be influenced by factors that relate to the subject–examiner dyad, the core of the traditional lie detection examination (Waid & Orne, 1981). For example, if the ethnic backgrounds of the two differ, autonomic arousal may increase in the subject. In addition, subjects tend to be more aroused when they are examined about crimes against a person as opposed to crimes against property (Waid & Orne, 1981). Hence, the impact of the examiner on the subject may vary in the traditional examination as a function of the type of crime under investigation. Given the dependence of the traditional examination on the subject–examiner interaction, there is remarkably little research assessing the variables important to it. A conservative approach necessitates minimizing this dependence.

No research has been done on ERPs that speaks to these issues. However, a large body of research in the psychological literature has documented experimenter effects on measures thought to be objective. In our view, therefore, the most effective ERP-based system will minimize interactions between the examiner and the subject, thereby reducing the possibility of any such confounding effects. There is no a priori reason to believe that the nature of the investigation and the type of stimulus presented should alter the properties of the ERP.

Note here that the nature of the crime and the details surrounding it are of fundamental importance to the development of a GKT. Essential elements of the crime, presumed to be known only by the guilty person, are used to develop the appropriate test items. These details must not be revealed to the public or to any suspects in the crime. Thus, a very thorough investigation of the criminal scene must be conducted, and the details of that investigation must be held in confidence. These demands place important limitations on implementing a GKT in field conditions, irrespective of the type of dependent that is used in the interrogation, and are probably the principal obstacles to broader use of this technique.

Countermeasures

There are a number of ways in which the autonomic measures used in the traditional polygraph examination can be altered to render them ineffectual. Subjects can produce a variety of relatively subtle movements (e.g., tongue biting, foot tensing); induce an altered cognitive state through drugs, hypnosis, or biofeedback; or do a secondary task concurrent with the polygraph test. A memory-impairing drug such as scopolamine can reduce the amplitude and prolong the latency of P300 (Meador, Loring, Adams, et al., 1987; Meador et al., 1989; Meador, Loring, Taylor, et al., 1987), and persistent blinking, mandibular jaw muscle contractions, or gross body movements can obscure the ERP. These variables must be considered in the development of any ERP-based procedure for detecting deception or inferring the possession of guilty knowledge. The traditional examination places the burden of identifying these countermeasures on the examiner. Proponents of the traditional lie detection procedures claim that it is highly improbable that countermeasures can succeed because the properly trained examiner would detect them (Reid & Inbau, 1977). Others have demonstrated, however, that these measures can be very effective at compromising the process (Honts, Hodes, & Raskin, 1985; Honts, Raskin, & Kircher, 1987; Kleinmuntz & Szucko, 1984; Lykken, 1981). Likewise, the use of appropriate countermeasures could obliterate distinctions between ERP responses to various test items.

Responses to relevant questions in the traditional polygraph test are enhanced among subjects who are attempting to deceive the examiner, independent of any overt behavior or the nature of the stimulus content, yet they are reduced in subjects who are less attentive to the relevant items (as measured by later memory; see review in Waid & Orne, 1981). The suggestion, then, is that as a countermeasure, the best strategy for the guilty subject may be to attempt to ignore the critical test items. Lykken's (1981) GKT is devised to counter this strategy by requiring the subject to repeat the response alternatives orally as they are presented, so that each is processed, at least minimally. An alternative strategy to counter an ERP-based interrogation that uses the oddball format would be to assign the critical items to the irrelevant category. Farwell and Donchin (1991) attempted to circumvent this problem by devising an oddball task whose intent was not revealed to the subject and by embedding the critical stimuli in the stimulus sequence as nontargets. In so doing, they demonstrated that it may be difficult to reassign the critical stimulus events and thereby influence measures of brain electrical activity. Their work suggests that the critical stimulus events may be categorized automatically by the subject and as such may be immune to attempts at recategorization. An alternative method is for the subject to do a secondary task (e.g., serial 7s—counting back from 100 in steps of 7) as the primary task, the polygraph test, is being done. We know from a solid behavioral and ERP literature that such dual-task performance is feasible and that P300 amplitude is reduced under these circumstances (Donchin, Kramer, & Wickens, 1986), the effect of which could be to obscure differences between relevant and other test items. Thus, sufficiently sensitive analytic techniques must be available to differentiate people who merely have small amplitude signals (and, consequently, small differences in brain responses to critical and noncritical test items) from those who engaged in a secondary task during the examination (and, consequently, also have small differences in the brain responses to critical and noncritical test items). The analytic procedures that have been used traditionally in ERP research may not be capable of making this distinction.

The Examiner

As we have discussed at some length throughout this review, the traditional polygraph examination demands much from the examiner. One such demand is that the examiner be skilled at convincing the subject of the infallibility of the process. If done successfully with an aroused innocent subject who is told that he or she looks deceptive on relevant questions (e.g., see case study in Lykken, 1981), his or her responses to critical items on subsequent tests may very well be enhanced. Indeed, from the influences of the variables we have just discussed on autonomic processes, it is straightforward to suggest that this scenario could easily explain the high incidence of false positives reported by many in traditional polygraphy. This suggests, of course, the importance of developing a procedure that neither depends on the communicational, observational, and intuitive skills of the examiner nor is based on questionable assumptions about the specificity of the psychophysiological measures of detection. Moreover, selection of dependent measures that are relatively unresponsive to the nonphysiologic variables that can confound autonomic responses is essential.

Concluding Thoughts

Development of procedures for inferring the possession of information that a person is unwilling to reveal offers a number of very complicated challenges, heretofore largely unappreciated by proponents of traditional polygraphy. Systematic scientific research is wanting in the area of traditional polygraphy, and the pursuit of such work may be frustrated by the type of psychophysiological measures used historically in this area. Some of these problems may be circumvented by using more direct measures of the cognitive processes that mediate the possession of guilty knowledge. ERPs hold forth such promise. However, they too can be influenced by many of the same variables that can contaminate autonomic measures. This must be acknowledged and investigated systematically. The P300 and N400 appear to be ideal candidates for use in this work because they can be elicited reliably under appropriate experimental conditions and a reasonably well articulated theoretical structure has evolved to explain their occurrence. Equally important, the experimental conditions in which these components are manifest can be adapted quite elegantly for implementation in circumstances that aim to infer the possession of critical pieces of information. Both components emerge under conditions in which a person distinguishes among stimulus events. They are manifestations of the neurocognitive processes engaged in making such distinctions. In the traditional P300 task, the experimenter defines the salient stimulus categories, and the cooperative subject sorts the stimuli accordingly. The recent work using P300 amplitude as the dependent measure in variants of the GKT suggests, however, that a person's categorization of stimulus events, at least into the relevant and irrelevant, may be engaged automatically and independent of the person's attempts at concealing this categorization. The conditions under which the N400 is elicited has similarly powerful implications for assessing the possession of guilty knowledge. First, the initial conditions under which it was observed required nothing of the subject other than to read the sentences as they were presented. Nonetheless, the N400 emerged to the semantically anomalous word. Hence, like the P300, the N400 may be engaged automatically under suitable experimental conditions. Second, it appears to be manifest under conditions in which a person is making an intentionally false response or is simply reading a false statement. Thus, the neurocognitive mechanisms indexed by the N400 may be engaged automatically when people recognize but are unwilling to reveal that they possess certain information or when they are processing information they know to be false, even if no overt response is required. These two components may provide very powerful indexes of the neurocognitive mechanisms that are activated by stimulus events of significance to a person, irrespective of the person's willingness to express that significance.

References

Allen, J. J., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences. *Psychophysiology, 29,* 504–522.

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

Balloun, K. D., & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. *Journal of Applied Psychology, 64,* 316–322.

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

Bashore, T. R. (1990). Age-related changes in mental processing speed as revealed in analyses of event-related brain potentials. In J. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-related potentials of the brain* (pp. 242–278). New York: Oxford University Press.

Bashore, T. R., & Osman, A. (1987, July). *On the temporal relation between perceptual analysis and response selection: A psychophysiological investigation of stimulus congruency and S-R compatibility effects on human information processing.* Poster presented at the Fourth International Congress of Cognitive Neuroscience, Dourdan, France.

Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M. (1982). An evaluation of polygraphers' judgments: A review from a decision theoretic perspective. *Journal of Applied Psychology, 67,* 701–713.

Ben-Shakhar, G., Lieblich, I., & Bar-Hillel, M. (1986). Trial by polygraph: Scientific and juridical issues in lie detection. *Behavioral Sciences and the Law, 4,* 459–479.

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology, 24,* 14–25.

Besson, M., Macar, F., & Pynte, J. (1984). Is N400 specifically related to the processing of semantic mismatch? *Society for Neuroscience Abstracts, 10,* Part 2, p. 847.

Bok, S. (1978). *Lying: Moral choice in public and private life.* New York: Pantheon Books.

Callaway, E. (1983). The pharmacology of human information processing. *Psychophysiology, 20,* 359–370.

Carroll, D. (1988). How accurate is polygraph lie detection? In A. Gale (Ed.), *The polygraph test: Lies, truth and science* (pp. 19–28). Newbury Park, CA: Sage.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 529–553.

Coles, M. G. H., Gratton, G., Kramer, A. F., & Miller, G. A. (1986). Principles of signal acquisition and analysis. In M. G. H. Coles, E.

Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes and applications* (pp. 183–226). New York: Guilford Press.

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? *Behavioral and Brain Sciences, 11*, 357–374.

Donchin, E., & Heffley, E. (1978). Multivariate analysis of event related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event related potential research* (EPA Publications No. 600/9-77-043, pp. 555–572). Washington, DC: U.S. Government Printing Office.

Donchin, E., & Herning, R. I. (1975). A simulation study of the efficacy of stepwise discriminant analysis in the detection and comparison of event-related potentials. *Electroencephalography and Clinical Neurophysiology, 38*, 51–68.

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. Porges (Eds.), *Psychophysiology: Systems, processes and applications* (pp. 244–267). New York: Guilford Press.

Donchin, E., Kramer, A. F., & Wickens, C. D. (1986). Applications of brain event-related potentials to problems in engineering psychology. In M. G. H. Coles, E. Donchin, & S. W. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 702–718). New York: Guilford Press.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenouse components of the ERP. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related potentials in man* (pp. 349–412). San Diego, CA: Academic Press.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology, 14*, 1–52.

Ekman, P. (1985). *Telling lies*. New York: Norton.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). 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.

Farwell, L. A. (1991, August 1). *Truth detector: A new paradigm for psychophysiological detection of concealed information ("lie detection")*. Technical report prepared for the U.S. Central Intelligence Agency.

Farwell, L. A., & Donchin, E. (1986). The "brain detector": P300 in the detection of deception [Abstract]. *Psychophysiology, 23*, 434.

Farwell, L. A., & Donchin, E. (1988a). Event-related brain potentials in interrogative polygraphy: Analysis using bootstrapping [Abstract]. *Psychophysiology, 25*, 445.

Farwell, L. A., & Donchin, E. (1988b, October). *The truth will out: Interrogative polygraphy with event-related brain potentials*. Poster presented at the 28th Annual Meeting of the Society for Psychophysiological Research,

Farwell, L. A., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology, 28*, 531–547.

Farwell, L. A., Martinerie, J. M., Bashore, T. R., Rapp, P. E., & Goddard, P. H. (in press). Optimal digital filters for long latency components of the event-related brain potential. *Psychophysiology*.

Fischler, I., Bloom, P. A., Childers, D. G., Arroyo, A. A., & Perry, N. W. (1984). Brain potentials during sentence verification: Late negativity and long-term memory strength. *Neuropsychologia, 22*, 559–568.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*, 400–409.

Fischler, I., Childers, D. G., Achariyapaopan, T., & Perry, N. W. (1985). Brain potentials during sentence verification: Automatic aspects of comprehension. *Biological Psychology, 21*, 83–106.

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, W. F., & Kopell, B. S. (1979). Event-related potentials recorded from young and old adults during a memory retrieval task. *Electroencephalography and Clinical Neurophysiology, 47*, 450–459.

Furedy, J. J. (1987). Evaluating polygraphy from a psychophysiological perspective: A specific-effects analysis. *Pavlovian Journal of Biological Science, 22*, 145–152.

Furedy, J. J., Davis, C., & Gurevich, M. (1988). Differentiation of deception as a psychological process: A psychophysiological approach. *Psychophysiology, 25*, 683–688.

Furedy, J. J., & Heslegrave, R. J. (1988). Validity of the lie detector: A psychophysiological perspective. *Criminal Justice and Behavior, 15*, 219–246.

Gale, A. (Ed.). (1988). *The polygraph test: Lies, truth and science*. Newbury Park, CA: Sage.

Giesen, M., & Rollison, M. A. (1980). Guilty knowledge versus innocent associations: Effects of trait anxiety and stimulus context on skin conductance. *Journal of Research in Personality, 14*, 1–11.

Goff, W. R., Allison, T., & Vaughan, H. G. (1978). The functional neuroanatomy of event-related potentials. In E. Callaway, P. Tueting, & S. H. Koslow (Eds.), *Event-related brain potentials in man* (pp. 1–80). San Diego, CA: Academic Press.

Gudjonsson, G. H. (1982). Some psychological determinants of electrodermal responses to deception. *Personality and Individual Differences, 3*, 381–391.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*, 66–85.

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70*, 177–187.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1985). Effects of socialization on the physiological detection of deception. *Journal of Research in Personality, 19*, 373–385.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology, 1*, 241–247.

Horneman, C. J., & O'Gorman, J. G. (1987). Individual differences in psychophysiological responsiveness in laboratory tests of deception. *Personality and Individual Differences, 8*, 321–330.

Horst, R. L., & Donchin, E. (1980). Beyond averaging: II. Single-trial classification of exogenous event-related potentials using stepwise discriminant analysis. *Electroencephalography and Clinical Neurophysiology, 48*, 113–126.

Horvath, F. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology, 62*, 127–136.

Iacono, W. G., Boisvenu, G. A., & Fleming, J. A. (1984). Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology, 69*, 289–299.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.

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

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. 2, pp. 69–138). Greenwich, CT: JAI Press.

Jutai, J. W., Hare, R. D., & Connolly, J. F. (1987). Psychopathy and event-related potentials (ERPs) associated with attention to speech stimuli. *Personality and Individual Differences, 8*, 175–184.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73*, 291–302.

Kleinmuntz, B., & Szucko, J. J. (1982). On the fallibility of lie detection. *Law and Society Review, 17*, 85–104.

Kleinmuntz, B., & Szucko, J. J. (1984). Lie detection in ancient and modern times: A call for contemporary scientific study. *American Psychologist, 39,* 766–776.

Kutas, M. (1985, February). *Event-related brain potentials in bilinguals during reading.* Paper presented at the 13th Annual Meeting of the International Neuropsychological Society, San Diego, CA.

Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11,* 99–116.

Kutas, M., & Hillyard, S. A. (1980b). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language, 11,* 354–373.

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

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11,* 539–550.

Kutas, M., & Hillyard, S. A. (1984a). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Kutas, M., & Hillyard, S. A. (1984b). Event-related brain potentials (ERP's) elicited by novel stimuli during sentence processing. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information: Event-related potentials* (pp. 236–241). New York: New York Academy of Sciences.

Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience, 1,* 38–49.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as an index of stimulus evaluation time. *Science, 197,* 792–795.

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 responses to semantic anomalies during reading, listening, and signing. In R. J. Ellingson, N. M. F. Murray, & A. M. Halliday (Eds.), The London symposia. *Electroencephalography and Clinical Neurophysiology, (Suppl. 39),* pp. 325–330. Elsevier Science: Limerick, Ireland.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of 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.

Lykken, D. T. (1959). The GSR in the detection of deception. *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. (1978). The psychopath and the lie detector. *Psychophysiology, 15,* 137–142.

Lykken, D. T. (1979). The detection of deception. *Psychological Bulletin, 86,* 47–53.

Lykken, D. T. (1981). *A tremor in the blood.* New York: McGraw-Hill.

Lykken, D. T. (1984). Trial by polygraph. *Behavioral Sciences and the Law, 2,* 75–92.

Lykken, D. T. (1985). The probity of the polygraph. In S. M. Kassin & L. S. Wrightsman (Eds.), *The psychology of evidence and trial procedure* (pp. 95–123). Beverly Hills, CA: Sage.

Lykken, D. T. (1988). Detection of guilty knowledge: A comment on Forman and McCauley. *Journal of Applied Psychology, 73,* 303–304.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21,* 171–186.

Marshall, N. K., Bashore, T. R., Miller, G. A., Coles, M. G. H., & Donchin, E. (1983). ERP consistency in young and elderly subjects [Abstract]. *Psychophysiology, 20,* 422.

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology, 59,* 477–488.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211,* 77–80.

McCauley, C., & Forman, R. F. (1988). A review of the Office of Technology Assessment report on polygraph validity. *Basic and Applied Social Psychology, 9,* 73–84.

Meador, K. J., Loring, D. W., Adams, R. J., Patel, B. R., Davis, H., & Hammond, E. J. (1987). Central cholinergic systems and the P3 evoked potential. *International Journal of Neuroscience, 33,* 199–206.

Meador, K. J., Loring, D. W., Davis, H. C., Sethi, K. D., Patel, B. R., Adams, R. J., & Hammond, E. J. (1989). Cholinergic and serotonergic effects on the P3 potential and recent memory. *Journal of Clinical and Experimental Neuropsychology, 11,* 252–260.

Meador, K. J., Loring, D. W., Taylor, H. S., Hughes, D. R., Gould, M. J., & Feldman, D. S. (1987). Neuropharmacology of the P3 potential. *Journal of Clinical and Experimental Neuropsychology, 9,* 49.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 27,* 3–67.

Mulder, G., Gloerich, A. B. M., Brookhuuis, K. A., van Dellen, H. J., & Mulder, L. J. M. (1984). Stage analysis of the reaction process using brain evoked potentials and reaction time. *Psychological Research, 46,* 15–32.

Neville, H. (1985). Biological constraints on semantic processing: A comparison of spoken and signed languages [Abstract]. *Psychophysiology, 22,* 576.

Ney, T. (1988). Expressing your emotions and controlling feelings. In A. Gale (Ed.), *The polygraph test: Lies, truth and science* (pp. 65–72). Newbury Park, CA: Sage.

Office of Technology Assessment. (1983). *Scientific validity of polygraph testing: A research review and evaluation—A technical memorandum* (Report No. OTA-TM-H-15). Washington, DC: U.S. Congress, Office of Technology Assessment.

Pfefferbaum, A., Ford, J. M., Johnson, R., Wenegrat, B. G., Kopell, B. S. (1983). Manipulation of P3 latency: Speed vs. accuracy instructions. *Electroencephalography and Clinical Neurophysiology, 55,* 188–197.

Picton, T. W., & Stuss, D. T. (1980). The component structure of the human event-related potentials. In H. H. Kornhuber & L. Deecke (Eds.), *Progress in brain research: Vol. 54. Motivation, motor and sensory processes of the brain* (pp. 17–49). Amsterdam: Elsevier/North-Holland.

Podlesny, J. A., & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology, 15,* 344–359.

Pritchard, W. (1981). The psychophysiology of P300. *Psychological Bulletin, 89,* 506–540.

Raine, A., & Venables, P. (1987). Contingent negative variation and P3 evoked potentials, and antisocial behavior. *Psychophysiology, 24,* 191–199.

Raine, A., & Venables, P. (1988). Enhanced P3 evoked potentials and longer P3 recovery times in psychopaths. *Psychophysiology, 25,* 30–38.

Rapp, P. E., Bashore, T. R., Martinerie, J. M., Albano, A. M., & Mees, A. I. (1989). Dynamics of brain electrical activity. *Brain Topography, 2,* 99–118.

Raskin, D. C. (1976). *Reliability of chart interpretations and sources of errors in polygraphy examinations* (Report 76-3, U.S. Department of Justice). Salt Lake City: University of Utah, Department of Psychology.

Raskin, D. C. (1978). Scientific assessment of the accuracy of detection of deception: A reply to Lykken. *Psychophysiology, 15,* 143–147.

Raskin, D. C., Barland, G. H., & Podlesny, J. A. (1978). *Validity and reliability of detection of deception.* (Final Report, 75-NI-99-0001, U.S. Department of Justice). Washington, DC: U.S. Government Printing Office.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15,* 126–136.

Raskin, D. C., & Podlesny, J. A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin, 86,* 54–59.

Reid, J. E., & Inbau, F. E. (1977). *Truth and deception: The polygraph ("lie detector") technique* (2nd ed.). Baltimore: Williams & Wilkins.

Ritter, W., Simson, R., & Vaughan, H. G. (1972). Association cortex potentials and reaction time in auditory discrimination. *Electroencephalography and Clinical Neurophysiology, 33,* 547–555.

Ritter, W., Vaughan, H. G., & Simson, R. (1983). On relating event-related potential components to stages of information processing. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research* (pp. 143–158). Amsterdam: North-Holland.

Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified event-related potential-based guilty knowledge test. *International Journal of Neuroscience, 42,* 157–161.

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.

Ruchkin, D., & Glaser, E. (1978). Simple digital filters for examining CNV and P300 on a single-trial basis. In D. A. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research* (pp. 579–581). Washington, DC: Environmental Protection Agency.

Ruchkin, D. S., Sutton, S., & Tueting, P. (1975). Emitted and evoked P300 potentials and variation in stimulus probability. *Psychophysiology, 12,* 591–595.

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy. *American Psychologist, 40,* 355–366.

Squires, K. C., & Donchin, E. (1976). Beyond averaging: The use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 41,* 449–459.

Strayer, D. L., Wickens, C. D., & Braune, R. (1987). Adult age differences in the speed and capacity of information processing: 2. An electrophysiological approach. *Psychology and Aging, 2,* 99–110.

Szucko, J. J., & Kleinmuntz, B. (1981). Statistical versus clinical lie detection. *American Psychologist, 36,* 488–496.

van der Molen, M. W., Bashore, T. R., Halliday, R. F., & Callaway, E. (1991). Chronopsychophysiology: Mental chronometry augmented by psychophysiological time markers. In J. R. Jennings & M. G. H. Coles (Eds.), *Handbook of cognitive psychophysiology* (pp. 9–178). New York: Wiley.

Waid, W. M., & Orne, M. T. (1981). Cognitive, social, and personality processes in the physiological detection of deception. *Advances in Experimental Social Psychology, 14,* 61–106.

Waid, W. M., Orne, M. T., & Wilson, S. K. (1979). Effects of level of socialization on electrodermal detection of deception. *Psychophysiology, 16,* 15–22.

Wood, C. C., McCarthy, G., Squires, N. K., Vaughan, H. G., Woods, D. L., & McCallum, W. C. (1984). Anatomical and physiological substrates of event-related potentials: Two case studies. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information processing: Event-related potentials* (pp. 681–721). New York: New York Academy of Sciences.

Working Group of the Scientific Affairs Board (1986). Report of the working group on the use of the polygraph in criminal investigation and personnel screening. *Bulletin of the British Psychological Society, 39,* 81–94.

Yohman, J. R. (1978). The guilty knowledge technique in lie detection. *Biological Psychology Bulletin, 5,* 96–103.

Received August 12, 1991
Revision received January 9, 1992
Accepted January 9, 1992 ■