Event-Related Potentials in Detection of Deception

J. Peter Rosenfeld
Department of Psychology
Northwestern University
Evanston, IL 60208

August 31, 1999

For Handbook of Polygraphy, ed. By Murray Kleiner, in press (2000), New York: Academic Press

Acknowledgement: I am profoundly grateful to the following colleagues, without whose creative participation and ideas, this work would not have been possible: Joel Ellwanger, Antoinette Reinhart Miller, Archana Rao, and Matt Soskins.

1. P300 and Event-Related Potentials

An event-related potential (ERP) is a series of peaks and troughs which appears in the Electroencephalogram (EEG) in response to occurrence of a discrete event, such as 1) presentation of a stimulus or 2) psychological reaction to a stimulus. The former ERP is called exogenous (or also an evoked potential), whereas the latter is called an endogenous ERP. ERPs are typically recorded from human scalps with standard EEG electrodes and amplifiers. Because the ERP is superimposed on the ongoing EEG, it is often difficult to see an ERP in a single sample. Typically, an eliciting event is repeatedly presented and the resulting ERPs (aligned at the time point of the eliciting event) are averaged to yield the subject's average ERP, in which activity not timelocked to the eliciting event averages out (i.e., to zero), and what remains is the timelocked wave series (see next paragraph, Fig. 1).

< insert Fig. 1 about here >

An endogenous ERP which has been extensively studied is the P300 (P3) wave. It is seen in response to rare, meaningful stimuli (often called "oddball" stimuli; Donchin & Coles, 1988; Johnson, 1988). For example, if a subject is viewing a Bernoulli (random) series of names, one every three seconds, and occasionally, one of these is the subject's name, a P3 wave is evoked in response to this rarely presented, recognized, meaningful (autobiographical) stimulus (see Fig. 1). P3 is a positive-going wave with a scalp amplitude distribution in which it is largest parietally (at Pz) and smallest frontally (Fz), taking intermediate values centrally (Cz). (Fz, Cz, and Pz are scalp sites along the midline of the head.) Its peak has a typical latency of 300-1000 msec from stimulus onset, and this latency varies with stimulus processing time, which is often determined by stimulus complexity. The size or amplitude of P3 at a given recording site is inversely proportional to the rareness of presentation; in practice, probabilities

≤ .3 are typically used. The meaningfulness of the stimulus is also influential in determining P3 size.

2. Early Studies of P3 in Deception Detection

The first published studies to utilize some of these attributes of P3 in detection of deception* were those of Rosenfeld and colleagues in the 1980s (e.g., Rosenfeld, Cantwell, Nasman, Wojdac, Ivanov, & Mazzeri, 1988). In these studies we utilized a guilty knowledge or concealed information test (CIT) paradigm with college student subjects, as follows: We instructed each subject to go to a private room to select an item from a box containing several small items such as a transistor radio, a camera, a wallet, a watch, and so on. The subject was instructed to remove the chosen item and pretend he/she was stealing it. The subject then had EEG electrodes attached and was seated before a display screen. On this screen, every few seconds, the name of an item would appear and elicit an ERP. One of the nine item names repeatedly presented at random was the name of the subject's chosen ("stolen") item. Separate ERP averages were collected by a computer for each of the nine presented items. Visual inspection of the ERPs allowed us to identify one and only one average ERP in each subject, which contained a distinct P3 component. Invariably it was the ERP in response to the chosen item. In these studies we also had an "innocent" control group whose members also looked at nine, repeatedly presented items, but none of these was the chosen item. In none of the 10 control subjects was a P3 detectable by visual inspection. An analysis of variance (ANOVA) on the group mean P3s showed a significantly larger P3 in the "guilty" than in the "innocent" group ($p<.001$). No individual diagnostics were done.

Group statistics have no use in the field, and in our follow-up studies of P3 in deception

*Other ERP components have been utilized also as reviewed by Rosenfeld (1995).

detection (e.g., Johnson & Rosenfeld, 1992), we tried to remedy the lack of individual diagnostics in the 1988 paper. We also used a somewhat different paradigm, one with features of a screening control question test (CQT), although it still had some CIT-like elements: Specifically, each subject was shown a list of antisocial or illegal acts not untypical of the college student population from which subjects were drawn: "Use false ID," "Cheat on test," "Plagiarize a paper," and so on. As subjects viewed these two or three-word phrases, their detailed meanings were spelled out on a standard audiotape, e.g., "'USE FALSE ID' means that sometime in the past five years, you used a document to show you were legally old enough to buy alcoholic beverages, when in fact you were not old enough and the document you used was false." From our first use of this paradigm (Rosenfeld, Angell, Johnson, & Qian, 1991), we knew that about 50% of our subject population would be guilty of "USE FALSE ID." The subjects were then wired for EEG recording as they heard and read the stimulus list, and we led them to believe that we were recording their physiology (we were not) as they experienced the stimulus list. This allowed us to execute the subsequent accusation phase of the study in which we told them that based on the previous "recording" phase, we suspected that they committed act A, but possibly also B, C, or D. Act B or C was always "USE FALSE ID," our relevant question (as in a CQT). The other three acts were always acts which we (but not the subjects) knew to have a low probability in the population. We designated B or C (whichever was unused as the relevant question) as the control question. We then seated subjects in the EEG recording chair, activated the computer and recording equipment, and randomly presented nine phrases (from the study list), one at a time, each repeated twelve times. We recorded the ERPs and collected average ERPs for each phrase. Two groups were formed naturally; a guilty group which had used false identification, and an innocent group which had not done so. We determined ground truth (who was really innocent and who guilty) by monitoring a checklist with a concealed, closed circuit TV camera. The checklist was completed by the subject after the ERP recording session, in a room which subjects believed afforded privacy. They were falsely told there would be one final brain wave test after we forced them to re-activate their memories by checking yes or no to the list of acts. Finally there was a final bogus recording session. Within each subject, the P3 amplitude (from one scalp site, Pz) to the relevant question was compared to that of the control question, of which all subjects were innocent. To make these comparisons, we utilized various methods, most notably, a bootstrap method (detailed later). There were no false positives in the innocent subjects and we correctly, diagnosed 77% of the guilty subjects. In an earlier study (Rosenfeld et al., 1991), we obtained similar results, but in that study there was a confound: The TV monitoring and checklist activity were done prior to the real ERP test sessions, allowing the interpretation that enhanced P3s to the relevant question were related to the fact that the relevant stimuli were made meaningful by their having uniquely been checked "yes," rather than their having been acts of which subjects were uniquely guilty. By running the ERP test prior to establishing ground truth, as in our later study, this confound was avoided.

These studies (certain details of which are here omitted for brevity) were designed to be analogues of the specialized version of the CQT used in employee screening. Thus, the questions were general, "did-you-do-it" type questions, and key comparisons were between responses to relevants and controls. On the other hand, the control questions were not developed on the basis of an interview, as they frequently are in field CQTs (which is a practice often criticized; e.g., Ben-Shakhar & Furedy, 1990). Our controls were false accusations, and so we were essentially comparing responses to real, correct information with responses to erroneous alternatives. This is a CIT-like element which avoids the typical criticism of the CQT (Ben-Shakhar & Furedy, 1990; Furedy, 1986).

It should also be noted that in addition to these studies just reviewed, there have been other reports of results with P3-based CITs, e.g., Farwell & Donchin (1991) and Allen, Iacono & Danielson, (1992). All this work was reviewed in Rosenfeld (1995) by way of criticizing an earlier review by Bashore & Rapp (1993).

3. The Simulated Malingering Paradigm

In the mid 1990s, we temporarily turned attention away from criminal and screening models, and began to utilize P3 to detect models of malingerers feigning cognitive deficit subsequent to mild head injury. As much of this work has just been reviewed (Rosenfeld & Ellwanger, in press), the present review will be brief.

The paradigms we have developed to detect malingered cognitive deficit are closely linked to data from the neuropsychological literature bearing on behavioral signs of malingering (Nies & Sweet, 1994). For one example, there are a small number of malingerers who claim to suffer loss of long term memories, such as autobiographical memories. It has been a straightforward matter to utilize P3 diagnostically in such situations. Ellwanger, Rosenfeld, Sweet, & Bhatt (1996) tested simulating malingerers stating (under instruction) that they could not recognize their birthdates presented in a Bernoulli series with other dates. This paradigm was very similar to the CIT described previously (Rosenfeld et al., 1988). There were distinct P3 responses uniquely to the rare (p=.12) birthday stimuli, but not to the frequent (p=.88) other dates in the simulators. Using a bootstrap test, 90% of the individual simulators were correctly detected.

The bootstrap test is utilized to demonstrate that the average P3 to the birthdate is larger than the average P3 to the other dates, despite the fact that the simulating malingerer is manipulated to verbally “mistakenly” misidentify about 50% of all stimuli. The idea is that if the brain signals that it recognizes the autobiographical stimulus (via the P3), the behavioral claim of non-recognition becomes difficult to sustain. One might expect that a simple t-test comparing the average P3 response of the birthdate with that of the response to other dates would be appropriate. The problem with this standard approach is that one would have to determine the average P3 within each compared category by first calculating the P3 response in each single sweep. Single ERP sweeps, as noted above, are very noisy, and we have shown previously (Rosenfeld et al., 1991) that t-tests based on single sweeps are relatively insensitive. The more accurate bootstrap approach works as follows:

Single sweeps are digitally filtered to pass low frequencies; 3 dB point: 4.23 Hz. P3 amplitude recorded at site Pz is used. The main question is whether , within each (lying or feigning) participant, the P3 amplitude is greater in response to a Birthday-oddball than to a frequent other date (Novel item). For each participant, a computer program draws at random *with replacement* a set of N1 (the total number of accepted Birthday-oddball sweeps for that participant) waveforms from the Birthday-oddball set of waveforms, and then averages this sample, and calculates P3 amplitude from this single (bootstrapped) average ERP. The program then draws randomly *with replacement* a set of N2 (the total number of accepted Novel-frequent sweeps for that participant) waveforms from the Novel-frequent set of waveforms, then calculates first an ERP average and then P3 amplitude from that average. The calculated Novel-frequent mean P3 is subtracted from the comparable Birthday-oddball value, and a difference value is obtained to place in a bootstrapped distribution which contains 100 values after 100 iterations of the process described above. In order to state with 95% confidence that Birthday-oddball and Novel-frequent P3s are indeed different, it is required that the value of zero or negative (see below) difference does not appear within 1.65 standard deviations from the mean of the bootstrapped distribution (see review by Wasserman & Bockenholt, 1989). A significantly larger amplitude in response to the Birthday-oddball items than that to the Novel-frequent items is then considered to be an indication of intact ability to recognize such autobiographical information. If this measure contradicts indications from the behavioral responses of the individual (which may give the impression of impairment) then the contradiction provides evidence for deceptive malingering of amnesia with respect to these memory items (Rosenfeld, Reinhart, Bhatt, Ellwanger, Gora, Sekera, & Sweet, 1998; Ellwanger, Rosenfeld, Sweet,& Bhatt, 1996). Note that the bootstrap approach is based on averages and avoids the problem of having to work with noisy single sweeps.

These bootstrap tests are always one-tailed. The difference values in the distribution are obtained by subtracting Novel from Birthday amplitudes, and a distribution containing only negative values would mean that Novel-frequent > Birthday-oddball in amplitude. (This outcome is the opposite of that signifying deception.) If zero is in the interval (lower end of the distribution is negative), the hypothesis that a larger P3 was elicited by the Birthday-oddball item in that participant is also rejected.

Many malingerers are too clever to deny recognition of basic autobiographical information. For such subjects, one needs a different approach: Thus, the other paradigm we have used to diagnose simulated cognitive deficit with P3 is a matching-to-sample procedure which is based on a similar, commercially available instrument lacking psychophysiological recording. This is basically an entrapment procedure in which an extremely simple test is given. Most normals and non-litigating head injury patients score > 90% correct on this simple test. It was thus suggested by some neuropsychologists that those test-takers scoring < 90% be classified as malingerers.

The problem, as we reviewed in Rosenfeld & Ellwanger, (in press), is that occasional head injury patients who are not litigating and who are likely doing their best on this test will honestly score well under 90% correct. We thus decided it would be helpful to enhance the simple behavioral test with P3 recording, as follows:

In this P3-based test, a sample 3-digit number is presented, then followed a few seconds later with a probe number which either perfectly matches or mismatches (in all 3 digit positions) the sample. Each trial consists of one sample followed by one probe. If the probability of a match trial is reduced below .2, a P3 will be evoked by a matching probe, but not by the other probes (mismatches). Using the bootstrap technique described above (in which we compare, within each subject, the P3 amplitude of the response to match with that of the mismatch), we correctly diagnosed up to 70% of the malingering simulators. In a more recently developed version of this approach, each sample was followed by nine probes, one of which was a perfect match. In this version of the procedure, 87% of subjects were correctly diagnosed (Rosenfeld & Ellwanger, in press; Rosenfeld, Ellwanger,Nolan, Wu, & Berman, 1999).

There are various control procedures we have utilized to address problems with the methods just reviewed. For example, if a malingerer is performing at or below 50% correct but his P3 to a match significantly exceeds the P3 to mismatches, then his claim of cognitive deficit is hard to make. However, if he is a clever malingerer, he could reason that he should perform above chance but below 90% correct, e.g., he should perform at about 75% correct. Thus, if his P3 to a match is greater than the P3 to the mismatch, his attorney could argue that 75% is still well below normal, yet does afford some match-mismatch discrimination which could account for the P3 difference. To deal with this problem in one way, we manipulated simulators to score at a dishonest 75% correct and found that these P3s to matches were larger than P3s to matches in a truth-telling group (Rosenfeld et al., 1998). (We had theoretical reasons to expect this as reviewed in Rosenfeld & Ellwanger, in press.) The point is that an attempt to foil the procedure by undershooting a target hit rate will result in a recognizably enhanced P3.

We also created a version of the match-to-sample test which was difficult (Ellwanger, Rosenfeld, Hankin, & Sweet, 1999). That is, on average, normal, truth-telling subjects doing their best would score at a legitimate 75% correct level. (What we did was to use seven and nine digit strings with mismatches off in only one digit position.) We did this so as to model honest patients with true deficit secondary to head injury, i.e., patients who might score only 75% on the easy (three digit) version of the test. In these cases, the match P3 was obliterated. There were good conceptual reasons to expect this effect also (Johnson, 1988, 1993; Rosenfeld & Ellwanger, in press). The point is that a genuine 75% hit rate is characterized by an absent P3, whereas a malingered 75% rate is characterized by an enhanced P3, relative to a normal subject telling the truth.

Finally, in connection with control procedures for autobiographical oddball paradigms, we ran a study (Ellwanger et al., 1999) in which we tested subjects using the experimenter's name as the oddball stimulus. Some subjects legitimately forgot it; others remembered, but were manipulated to simulate forgetting. The real forgetters showed no P3s, the simulators showed large P3s. We conclude that P3 amplitude is a reliable sign of recognition or the lack thereof despite deceptive verbal behavior.

4. <u>P3 Profile as a Novel Dependent Index</u>

To this point, I have considered the amplitude of P3 as an index of recognition in various paradigms. The amplitude has been typically measured in these studies, at one scalp site, Pz, where P3 is usually maximum. I also stated above that P3 amplitude is typically largest parietally (Pz) and the smallest frontally (Fz). However there are many ways that such a distribution may be achieved (see Fig 2). Moreover, there are circumstances which can alter even the basic pattern of Pz > Cz > Fz. Johnson (1988,1993) has given elegant theoretical accounts of the significance of varying scalp profiles in which he emphasises that each particular profile represents a unique pattern of activation of P3 neurogenerating neurons associated with a particular psychological state or task. Thus if two tasks or conditions, within or between subjects, produce differing amplitude distributions (or profiles), one may infer that differentially located groups of P3-neurogenerating neurons are involved by the two conditions. Although there are other explanations for differing profiles (e.g. Donchin, Spencer, & Dien, 1997), all are consistent in assuming that differing profiles in two states means that the brain is working differently in the two states. The profiles are actually scaled scalp distributions, the scaling being necessary to guarantee that the scaled amplitude profiles are orthogonal to simple amplitude differences (McCarthy & Wood, 1985; Johnson, 1988,1993). The typical familiar statistical method of showing that two group profiles differ at the same three sites is to do a 2 (groups) X 3 (sites) ANOVA on the scaled amplitudes and show a significant group by-site interaction.

<insert Fig. 2 about here>

We have thus recently utilized the scalp profile as another brain wave indicator of the two states of deception versus truth-telling. The results of some of these studies have not only yielded a new direction for P3-based detection of deception, but have in fact pointed to the possibility of our someday observing a specific lie response (which Lykken called "a dream" in Lykken, 1981). However, there are some caveats about this possibility which are discussed later.

One of our first studies (Rosenfeld et al. 1999) along this line involved the use of our match-to-sample paradigm with nine probes. There were two groups of subjects: liars (L) and truth-tellers (T). The T-subjects were told to do their best on the easy test, and that they would probably score 100% correct. The L-subjects were manipulated to score 50% correct on both matches (MAT) and mismatches (MIS). Thus for T-subjects there were two possible stimulus-responses combinations: 1) Match stimulus and Match response (Mat-Mat), 2) Mismatch stimulus and Mismatch response (Mis-Mis). For the L-subjects there were two additional possible combinations, Mat-Mis and Mis-Mat on dishonest trials. One major finding was that a comparison of T and L subjects' scaled P3 amplitudes as a function of site (Fz, Pz, Cz) and stimulus type (regardless of response) yielded a significant group-by-site interaction, meaning that the P3 profiles of the truth group differed from those of the liar group. This is seen in Fig. 3 where both the Mat and especially the Mis profiles of the liar group (called “malinger” in the figure) show a quadratic component, whereas the T-profiles appear more linear (or, as we jest in the lab , "liars are crooked.")

<insert Fig. 3 about here>

The other major finding of this study is seen in Fig. 4. There, the four scaled P3 profiles associated with all four stimulus-response combinations in the liar group are shown. What is noteworthy is that within just the liars, regardless of the stimulus type, the profiles superimpose, suggesting a deception-specific profile. In the truthtellers, the two corresponding profiles clearly differ, as they should, since the brain probably does process matches differently than mismatches and thus the associated neurogenerator sets recruited by the two kinds of processing should be different. However, engaging in deception appears to swamp out these effects, as just noted.

(All the results just presented were statistically confirmed.) As we will see, however, in other experiments, this swamping out does not happen .

<insert Fig. 4 about here>

As the results just described were being collected , we were prompted to re-analyze profile data from some older published studies in which we had the profile data set but analyzed it only for simple amplitude and latency effects at one site only. One study utilized the autobiographical oddball paradigm with subjects' birthdates as oddballs. The other study utilized the original match-to-sample paradigm with one test probe per sample. In each study, there were two conditions, a truth-telling condition and a lie condition in which subjects were instructed to lie on about 50% of the trials. It is well worth noting that these studies were done two years apart by two experimenters on two sets of subjects. The key results (from Rosenfeld et al., 1998) are seen in Fig. 5. What are shown are scaled P3

<insert Fig. 5 about here>

profiles from Fz, Cz, Pz for the liars and truth tellers in both experiments. If one looks only at the truth conditions, there is a clear interaction (statistically confirmed) between the autobiographical and match-to-sample studies. This was quite expectable since the two paradigms have obvious differences, requiring differential cognitive processes which should activate differentially located sets of neurogenerator neurons, a situation resulting (see above) in differing scalp profiles. In contrast, the profiles from the two paradigms in the lie conditions yielded no significant interaction and are indeed virtually superimposed between Cz and Pz. Again, there seems to be a deception-specific profile which seems to swamp out other influences which can express their effects in truth-tellers. This is comparable to the situation in Fig. 4, discussed above.

In another study (Miller, 1999a) using an autobiographical paradigm, two groups of subjects, truthtellers (T) and liars (L), were run in two blocks each. In the first, the autobiographical oddball stimulus was the subject's phone number, and all subjects were instructed to respond truthfully. In the second block where the oddball was the birthdate, the T-subjects responded truthfully, whereas the L-group lied on 50% of the trials. Fig. 6 below (from Miller, 1999a) shows that only liars ("malingerers" in the figure) have a P3 profile which differs from the others (all truth-telling blocks), an effect which was statistically supported. When the

<insert Fig. 6 about here>

lie and truth response trials are separately plotted for the L-group as in Fig. 7, however (from Miller 1999a), no interaction appears nor is found statistically. Again, it appeared that the deceptive mind-set carries over in truth-telling trials, swamping out the effect of specific behavioral response.

<insert Fig. 7 about here>

These kinds of datasets generate hope that a specific lie response (a deception-specific profile) can be found. Unfortunately, things are not so simple. Another study (Miller, 1999b) used exactly the same methods as the one just described, with the following differences:
1) There were two groups, both of which told the truth in Block 1 but lied in Block 2. These two groups were selected to be either high or low in psychopathic attributes. 2) Seven scalp sites were used (F3, F4, Fz, Cz, P3, P4, Pz) instead of the usual midline three used in other studies heretofore described. 3) A different experimenter ran the subjects.

The profile results, to the extent that they are comparable, were opposite to what was obtained in Miller (1999a). Fig. 8 below which is comparable with Fig. 6, shows no interactions (and none were statistically detected); the subjects' profiles did not differ between deceptive and truthtelling blocks (regardless of high or low psychopathy, a variable which had no effects on

<insert Fig. 8 about here>

scaled P3 profile in these studies as the figure shows). On the other hand, Fig. 9 shows a statistically confirmed interaction between deceptive vs honest response trials and scalp site. Here, for the first time it has been noted in this review, within one block, deceptive profiles differed from truthful profiles in the same subjects. (Again, psychopathy level had no effect). Why then were there no interactions in Fig. 8? Perhaps because the truth trials diluted the effect of deception. Why this occurred in Miller (1999b) but not Miller (1999a) is not clear, but probably relates to the differences between the two studies listed above. The point is that profile effects are likely influenced by many cognitive and emotional variables in addition to deception, and any putative profile-based specific lie responses found in the future may not generalize beyond the paradigm utilized. (More evidence on this point will be presented shortly. We wish to first note here parenthetically that using specialized bootstrap tests on amplitudes only at Pz, as in our earlier studies, 100% of the low psychopathy individuals were detected (12/12) and 92% (12/13) of the high psychopathy individuals were detected during deception blocks. There was no difference, statistically, between these proportions.)

<insert Fig. 9 about here>

There are other inconsistencies across our studies which temper expectation of finding a generalized specific lie response. For one thing it can be seen in this review that liars in some figures have upward concavities in profiles whereas in other figures, the concavities are downward. Associated with these effects, is another inconsistency. In the scaled profiles of Rosenfeld et al., 1998 (autobiographical study, Fig. 5) the scaled truth condition Fz oddball value was greater than the comparable value in the liar group. In contrast, in Fig. 6 above, the liar group's lie block profile shows an enhanced Fz value in comparison with the truthful profiles. In the match-to-sample paradigm of Rosenfeld et al., (1998, Fig. 5) the same pattern is seen. These contrasts (all based on standard base-to-peak measurement of P3*) could be related to paradigm and experimenter differences, to slight changes in instructions to subjects, to between versus within group comparisons, and so on. They reinforce the point that deceptive profile effects will probably not be straightforward, but will be complex, interacting with other paradigm-elicited cognitive and emotional variables, and thus will require much tuning before they can be made field-ready. However it must be emphasized that in several studies to date, we do consistently find differences between liar and truthteller profiles. This suggests that once paradigms are standardized, the basic method may be utilized in the field. The next study to be described (the last in this section) also failed to find swamping effects (Fig. 10) and like Miller (1999b, Fig. 6) illustrates that only liars when they lie produce an aberrant profile. The study was also undertaken for more theoretical reasons, however.

<insert Fig. 10 about here>

The aim of Rosenfeld, Rao, Soskins, & Miller (submitted) was mostly theoretical. For practical detection of deception, it does not matter why the truth-teller profile is different than that of the liar. This difference <u>may</u> relate to a specific deception signature in the P3 profile, but it need not have to be so. It may well be that the previously described truth-teller profiles differ

---

*In many of our earliest studies, we used a peak-to-peak method of P300 determination. We still think it best for diagnostic bootstrap tests of simple amplitude at one site. However, in profile work, for both theoretical and empirical reasons, we use the standard base-to-peak index, although in one of our earliest studies (cited here as Rosenfeld et al., 1999) we used both methods, and Fig. 4 here is based on peak-to-peak values. The comparisons made across studies in this paper, with the exception of Rosenfeld et al., (1999), all used the base-to-peak method.

from associated liar profiles for the simple reason that our deception instructions place a greater task demand on the liar, who must try to track a random appearing, target lie rate of 50% and thus decide on (or just prior to) each trial whether or not to lie. In contrast, the truth-tellers have none of this work to do, they simply tell the truth. This probably accounts for our typical finding that simple unscaled P3 amplitude is typically greater in the truthteller, than in the liar. Doing multiple tasks along with an oddball-P3 task typically reduces P3 amplitude; (Kramer, Sirvaag, & Braun, 1987). Of course the scaled P3 profile is independent of simple amplitude effects; that is the point of scaling, so truth-teller/liar scaled profile differences are not necessarily related to simple amplitude effects.

In Rosenfeld et al., (submitted), we tried to create a control task in which no deception would be occuring, but in which control subjects would have task demands approximating those of liar subjects. Again we used a paradigm similar to that in Miller (1999a,b) with two blocks of trials, a truth-telling block for all subjects, followed by a second block in which liars were told to lie on about half the trials to both rare autobiographical and frequent non- autobiographical stimuli. The control group, on the second block, was instructed to respond truthfully, but to then repeat aloud backwards about half the stimuli, (a random half). To equalize the speaking aloud demand, the liars were told to repeat aloud all stimuli normally (forwards) after giving the yes or no response (birthdate or not). Fig. 10 shows scaled profile results for oddball (autobiographical) stimuli in both blocks, for both groups, and for both response types (yes/no) in each of the groups. What is dramatically apparent is the fact that all profiles seem alike except that of liars on the second block during lie trials, and this visual impression was statistically confirmed. This result supports the existence of a specific lie response, at least in this paradigm, and suggests in two ways that the difference between the lie profile and the other profiles is not simply a matter of difference in task demand: 1) The truth-telling, control group has equivalent task demands but a different profile than that of the liars; 2) within the liar group in the second block, there is a different profile associated with lie and truth trials. It is clear that within the lie group, task demand must be constant from trial to trial. One could argue that the difference between results with the control group and those of the liar group was due to the difference in cognitive processing elicited by the two tasks (backwards repetition versus deception). This is not consistent however with the similarity (Fig. 10) of forwards and backwards response profiles in the control group on the second (test) block. However, the second point just noted provides incontrovertable evidence that task demand is not the source of the difference between liars lying and liars telling the truth all within their second block.

This dramatic difference between the profiles of the liars lying versus liars on truth-telling trials is quite consistent with the results of Miller (1999b). In that study also, there is no swamping on truth trials of effects of the liar mindset on lie trials within the second (test) block. These results suggest a specific lie response at least within the paradigm. However, as noted above, we do not routinely obtain this result. The swamping by deception effects across experiments and paradigms (Rosenfeld et al., 1998) noted above, is not the same process, it must be recalled, as the (lack of) swamping across trials within a time block. The latter did occur in Miller (1999a), a comparable autobiographical oddball paradigm (unlike Rosenfeld et al., 1999, Fig. 4, which used a 9-probe, match-to-sample paradigm). We don’t know what the within block response type effect was in the also comparable autobiographical paradigm of Rosenfeld et al., (1998), because at that time we lacked the software to obtain it. Thus there are three studies in which the liar groups are comparable: Miller (1999a), Rosenfeld et al., (submitted) and Miller

(1999b). Of these three, the latter two show distinct (lie versus truth-telling) response type distributions supporting the notion of specific lie profiles within a paradigm.

It must be emphasized that most of the studies reviewed in this section utilized just three scalp sites, an electrode density which is low by today's standards where 20-30 leads are common and 64-256 electrode studies are becoming less unusual. I am of the view that as our electrode density increases (we are presently recording from 30 channels), greater profile specificities will be seen. In this connection, it should be recalled that in Miller (1999a) which did not find response specificities, only three sites were utilized, but in Miller (1999b) which used seven electrodes, the response specificities were obtained. It will be recalled that these two studies were otherwise very similar, paradigmatically.

One issue we have not yet addressed with regard to profile analysis is the matter of intraindividual diagnostics. The bootstrapping methods described above are obviously not appropriate for profile comparisons, as they involve amplitude comparisons at one site. At the time of writing of this chapter, we are in the process of developing such methods. Since we do not now know which will work, and which will not, we will refrain from presenting our as yet untried diagnostic algorithms.

5. False Memory Studies

False memory refers to honestly believed recall of events which did not happen (Miller, 1999c). This is clearly a highly topical subject in the United States, where it has involved claims made by psychiatric outpatients of childhood abuse (Loftus & Pickrel, 1995; Loftus, 1997). The claims have been typically stimulated by apparent retrieval of memories in the course of psychotherapy. Some of these claims turn out to be true, but many have been later recanted as in error. When the claims turn out to be false, great and irreparable damage can be done to those wrongly accused. When the claim is malingered (i.e., the claimant knows full well of the innocence of the accused) damage to accusees is criminal. It is also likely that some honestly believed memories are never recanted although they are false. In this case, there is no remedy for falsely accused innocents. In any of these situations, it would be well to have an objective method of discriminating real and false memories of whatever subtype.

Honestly believed false memories have been modeled in the laboratory utilizing a method developed by Deese (1959) and Roediger & McDermott (1995). In this paradigm, a subject studies a list of words (OLD words) which are all associated with a word (LURE) which is not presented in the study session. For example, a subject might study words such as stars, bed, dream, pajamas, darkness and so on, all associates of the word night which is not presented for study. In a test session following the study session, the subject is tested on OLD words, LURE words, and NEW words, which are words which were not presented in the study session and which are not associatively related to OLD words (as the LURE words are). The subject's task is to signal "OLD" if he/she recognizes the word. Otherwise, the response is to be "NEW." There is much evidence (reviewed by Miller, 1999c) that subjects will genuinely believe that they recognize 50-80% of the LUREs, as they do for 99% of the OLDs. (Recognized LUREs are here referred to as LURES-OLD. Correctly rejected LUREs are called LURES-NEW).

Miller (1999c) utilized a variant of this paradigm in our lab in which the probability of LUREs and OLDs in the test session was .15 each, leaving the probability of NEW words at .7. Miller recorded ERPs during word presentation in the test session and expected that rare OLD and LURE words would evoke the P3. We further expected that the P3 scaled profiles would differ between OLD and LURE trials as it seemed that the brain would process these two stimulus types in a different manner, and thus, we would have a physiological means of distinguishing real and false memories.

We were disappointed: Fig. 11 shows that OLD and LURE-OLD scaled profiles at seven scalp locations seem to superimpose and indeed, no word-type by site interaction was found. (Fig. 11 shows peak-to-peak values. However, no word-type-by-site interaction was found with base-peak values, either.) Similarly, P3s evoked by NEW and LURE-NEW had overlapping

<insert Fig. 11 about here>

distributions. When we combined OLD and LURE-OLD distributions, and also combined NEW and LURE-NEW profiles, we did obtain a significant response type ("OLD" or "NEW") -by -site interaction, meaning that the cognitive process leading to the differing responses involve differing neurogenerator sets. However, we were still left without a way of distinguishing real and false memories--until we looked at the P3 latencies (times from stimulus presentation to P3 peak). Fig. 12 (from Miller, 1999c) shows, remarkably, that the LURE-OLD latency is dramatically less than the OLD latency ($P<.03$). This result had important implications, and so the experiment was entirely replicated by another experimenter on another group of subjects. The results also largely replicated, in particular, the latency results ($p<.006$). The possible reason for this effect (and the fact that the LURE-OLD latency is the shortest seen in Fig. 12) is well-described by Miller (1999c), and space limitations prevent restating Miller's hypothesis here. The significance of the results on latency is that they suggest that P3 latency is a correlate of the brain's unconscious recognition of false memory, even as the subjects genuinely and strongly believe in their memory illusions. There is the further implication that we may, in P3 latency, have after all a potentially practical means of distinguishing real and false memory.

<insert Fig. 12 about here>

The main obstacle in the way of realizing this goal is the difficulty of doing a diagnostic analysis of P3 latency within an individual. This is because of the variability of latency over trials. The bootstrapping technique used within individuals as described above for use with P3 amplitude has been tried also with latency in these false memory studies and has a sensitivity just greater than .5, which is unacceptable . The problem is that calculating a P3 latency in a bootstrapped ERP average will frequently involve a P3 wave which is very broad, due to the smearing of P3 waves of differing latencies over trials. Alternatively, there will be two or three peaks in the bootstrapped average ERP in the time window in which we search for P3. Our current algorithm for selecting latency cannot accurately determine P3 latency in these cases. It simply looks for the maximum positive peak in a window usually extending from 300 to 1500 ms. If there is a broad peak or multiple peaks, two results often seen in this false memory paradigm, the algorithm may select a non-representative latency. We are currently working on new algorithms for latency diagnostics within individuals.

6. Summary

Based on several studies , I believe that the P3 component of the ERP, which is clearly related to cognitive processes, has great potential in detection of deception. Its amplitude, profile, and latency have all been seen to correlate with deception-related phenomena. As of this writing, however, there is a great deal of work to be done in 1) paradigm development and selection, which will depend on the needs of potential users and 2) tuning of intraindividual analysis methods, particularly for profile and latency.

References

Allen, J., Iacono, W.G. and 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.

Bashore, T.T., & Rapp, P.E. (1993). Are there alternatives to traditional polygraph procedures? *Psychological Bulletin, 113,* 3-22.

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

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58(1),* 17-22.

Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral Brain Sciences, 11,* 357-374.

Donchin, E., Spencer, K., & Dien (1997). The varieties of deviant experience: ERP mainfestation of deviance processors in Van Boxtel, G.J.M. & Bocken, K.B.E. (Eds.), *Brain and Behavior: Past, Present, and Future,* Tilburg Univeristy Press, p. 116.

Ellwanger, J., Rosenfeld, J.P., Sweet, J.J. & Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *International Journal of Psychophysiology, 23,* 9-23.

Ellwanger, J., Rosenfeld, J.P., Hannkin, L.B., & Sweet, J.J. (1999). P300 as an index of recognition in a standard and difficult match-to-sample test: A model of Amnesia in normal adults. *The Clinical Neuropsychologist, 13,* 100-108.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). The definition, identification, and reliability of measurement of the P3 component of the event related brain potential. In P.K. Ackles, J.R. Jennings, & M.G.H. Coles (Eds.), *Advances in psychophysiology, vol. 2.* Greenwich: JAI Press.

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

Furedy, J.L. (1986). Lie detection as psychophysiological differentiation. In G. Coles, M. Donchin, & S. Porges (Eds.), *Psychophysiology: System, processes, and applications* (pp. 683-701). New York: Guilford Press.

Johnson, M.M. & Rosenfeld, J.P. (1992). Oddball-evoked P300-based method of deception detection in the laboratory II: Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology, 12*, 289-306.

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

Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology, 30*, 90-97.

Loftus, E.F. (1997). Memory for a past that never was. *Current Directions in Psychological Science, 6*, 60-64.

Loftus, E. & Pickrell, J.E. (1995). The information of false memories. *Psychiatric Annals, 25*, 720-725.

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

McCarthy, G. & Wood, C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroenceph. Clin. Neurophysiol., 62,* 203-208.

Miller, A.R. (1999a). P300 amplitude and topography in pseudomemory phenomena. Unpublished Doctoral Dissertation, Northwestern University, Evanston, ILL., pp. 11-59.

Miller, A.R. (1999b). P300 amplitude and topography in pseudomemory phenomena. Unpublished Doctoral Dissertation, Northwestern University, Evanston, ILL., pp. 60-122.

Miller, A.R. (1999c). P300 amplitude and topography in pseudomemory phenomena. Unpublished Doctoral Dissertation, Northwestern University, Evanston, ILL., pp. 123-144.

Nies, K., & Sweet, J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9,* 501-552.

Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(4)*, 803-814.

Rosenfeld, J.P. (1995). Alternative views of Bashore and Rapp's (1993) alternatives to traditional polygraphy: A critique. *Psychological Bulletin, 117*, 159-166.

Rosenfeld, J.P., Angell, A., Johnson, M., & Qian, J. (1991). An ERP-based, control-question lie detector analog: Algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology, 38*, 319-335.

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, 24*, 157-161.

Rosenfeld, J.P. & Ellwanger, J.W. (in press). Cognitive psychophysiology in detection of malingered cognitive deficit. In J.J. Sweet, ed., *Forensic Neuropsychology: Fundamentals and Practice*, Lisse, Netherlands: Swets & Zerlanger Publishers.

Rosenfeld, J.P., Ellwanger, J.W., Nolan, K.,Wu, S., & Bermann, R. (1999). P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model. *Int. J. Psychophysiol.,33(1)*, 3-20.

Rosenfeld, J.P., Reinhart, A.M., Bhatt, M., Ellwanger, J., Gora, K., Sekera, M., & Sweet, J. (1998). P300 correlates of simulated malingered amnesia in a matching-to-sample task: topographic analyses of deceptive vs. truthtelling responses. *International Journal of Psychophysiology, 28(3)*, 233-249.

Rosenfeld, J.P., Sweet, J.J., Chuang, J., Ellwanger, J. & Song, L. (1996). Detection of simulated malingering using forced choice recognition enhanced with event-related potential recording. *The Clinical Neuropsychologist, 10(2)*, 163-173.

Rosenfeld, J.P., Rao, A., Soskins, & Miller (submitted). Scaled P300 scalp distribution correlates of deception in an autobiographical oddball paradigm.

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

FIGURE CAPTIONS

Fig. 1. Average ERP responses at scalp sites Fz, Cz, Pz, and EOG, to a subject's name presented with probability = .12, darker curve, and to other dates, presented with probability (p) = .88. Note (arrow in Pz waves, lower left) the prominent P300 response in the ERP in response to the rarer, more meaningful stimulus, but absent in the other waveform (Positive is down, negative is up in all ERP figures). EOG is electrooculogram, recording of eye movements whose flatness indicates freedom from eye movement artifacts.

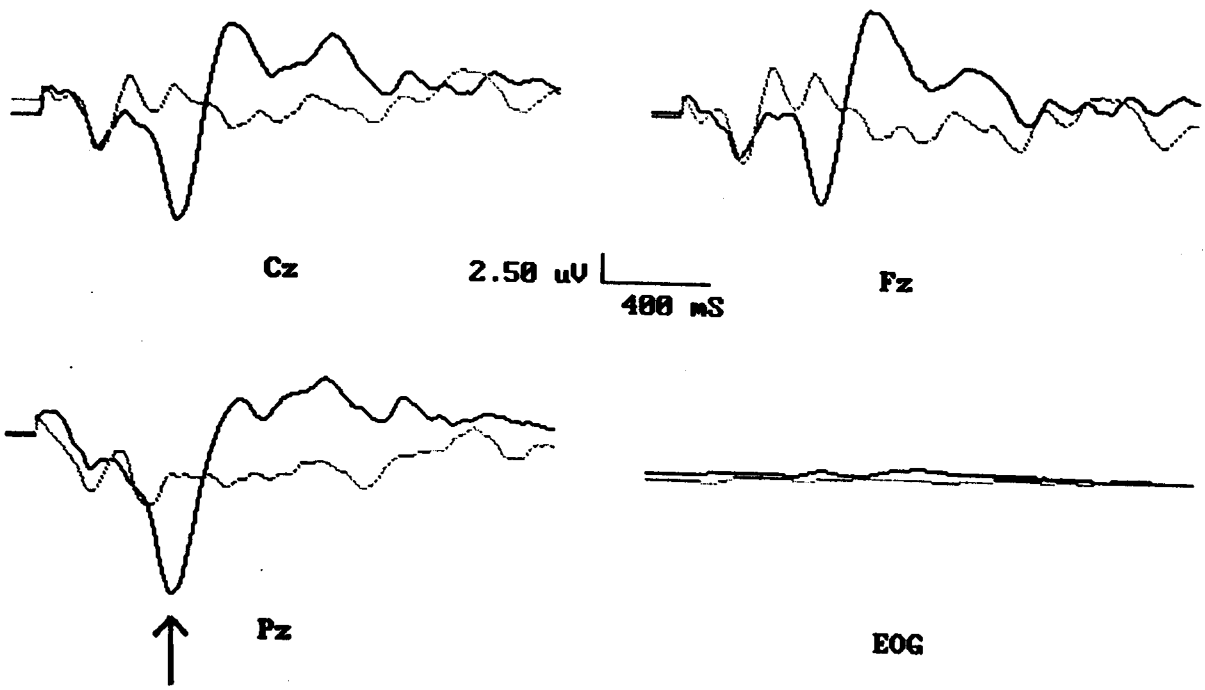

Fig. 2. Hypothetical plot of scaled P3 amplitude at 3 scalp sites under 5 different task conditions. In all cases, Pz (site 3) > Cz (site 2) > Fz (site 3), yet not all curves are parallel.

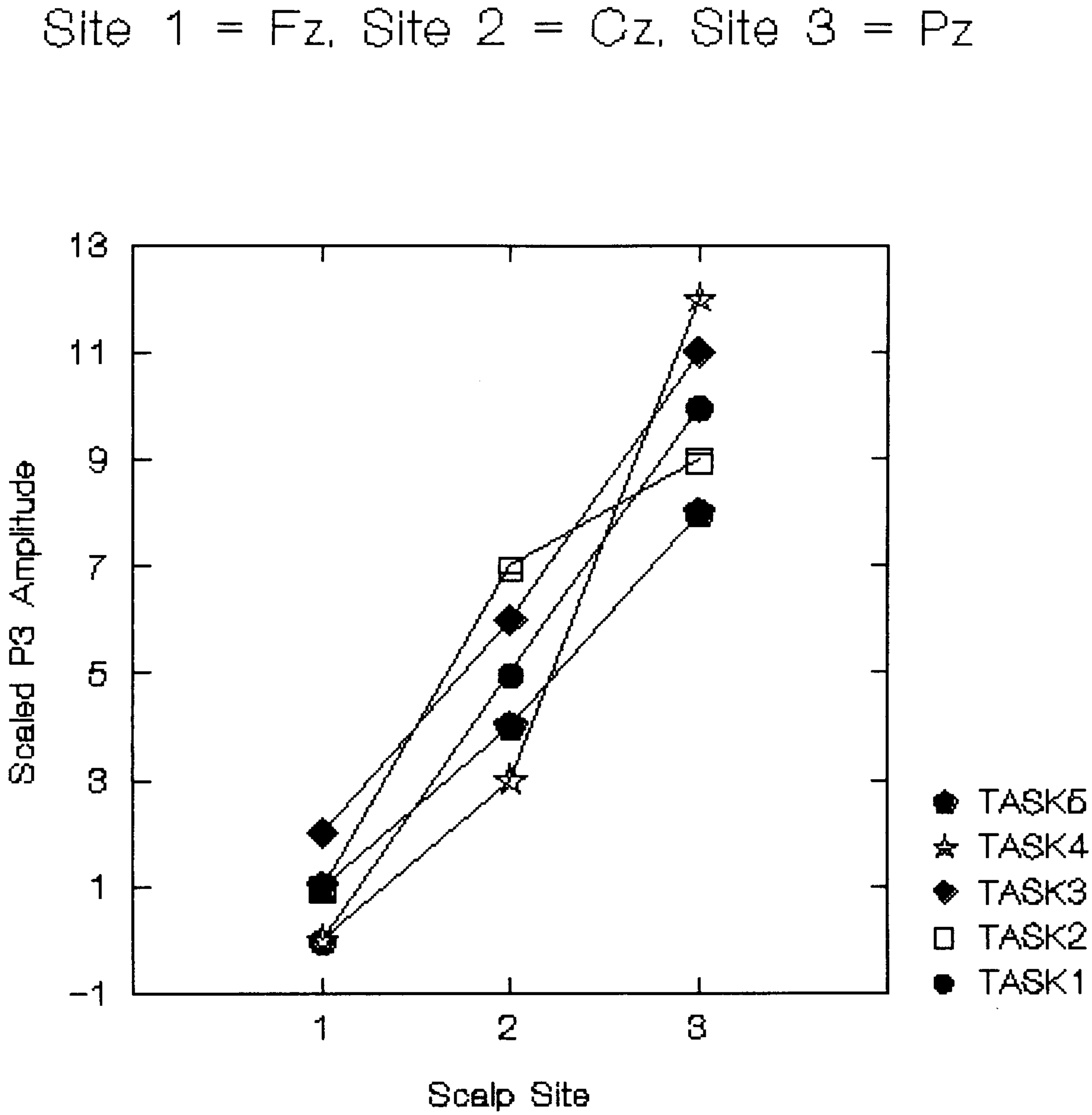

Fig. 3. Scaled P3 amplitude distributions for truth and liar (called "MALINGERER" in figure) groups associated with the four stimulus-response combinations shown.

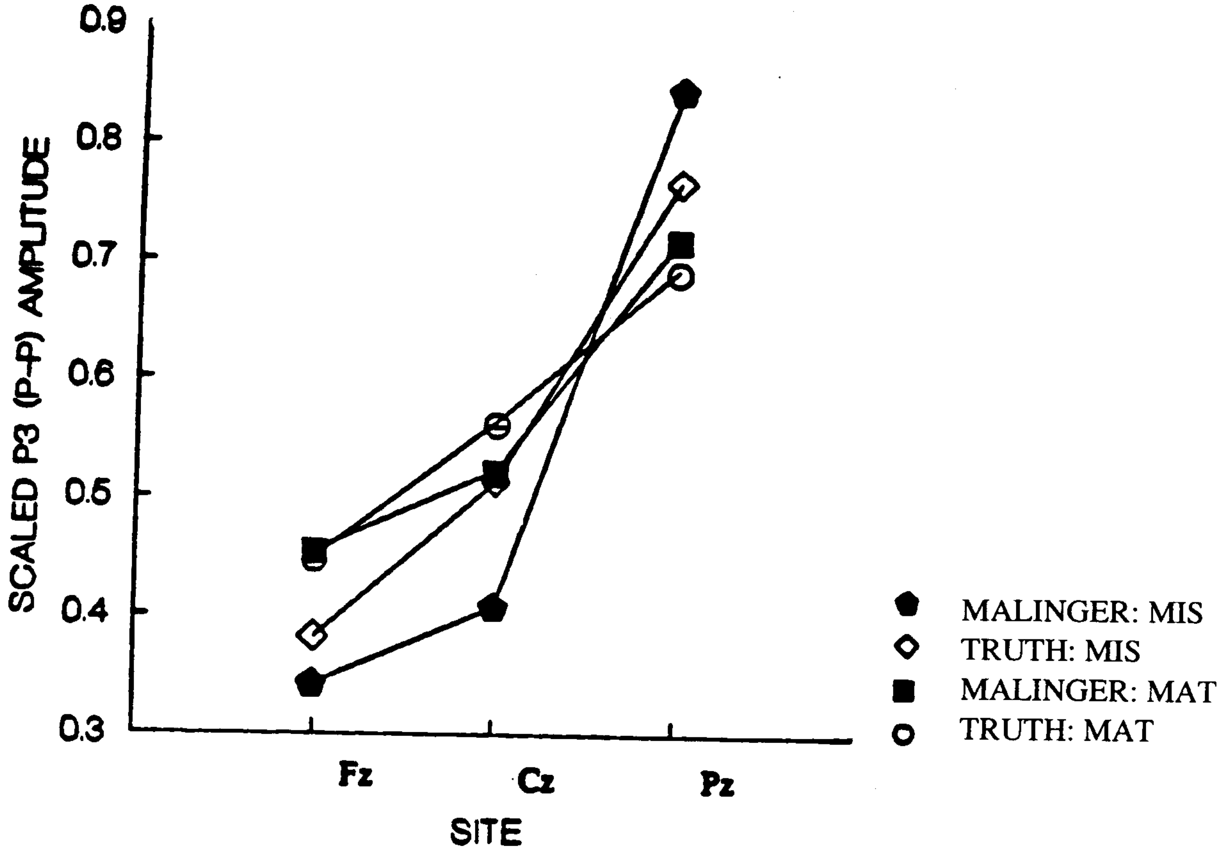

Fig. 4. Computer calculated, scaled P300 amplitudes as a function of site within liar group during 4 types of stimulus-response combination: match-truthful (MAT-MAT) match-deceptive (MAT-MIS), mismatch-truthful (MIS-MIS), and mismatch-deceptive (MIS-MAT). Note the superimposability of deceptive responses, regardless of stimulus type.

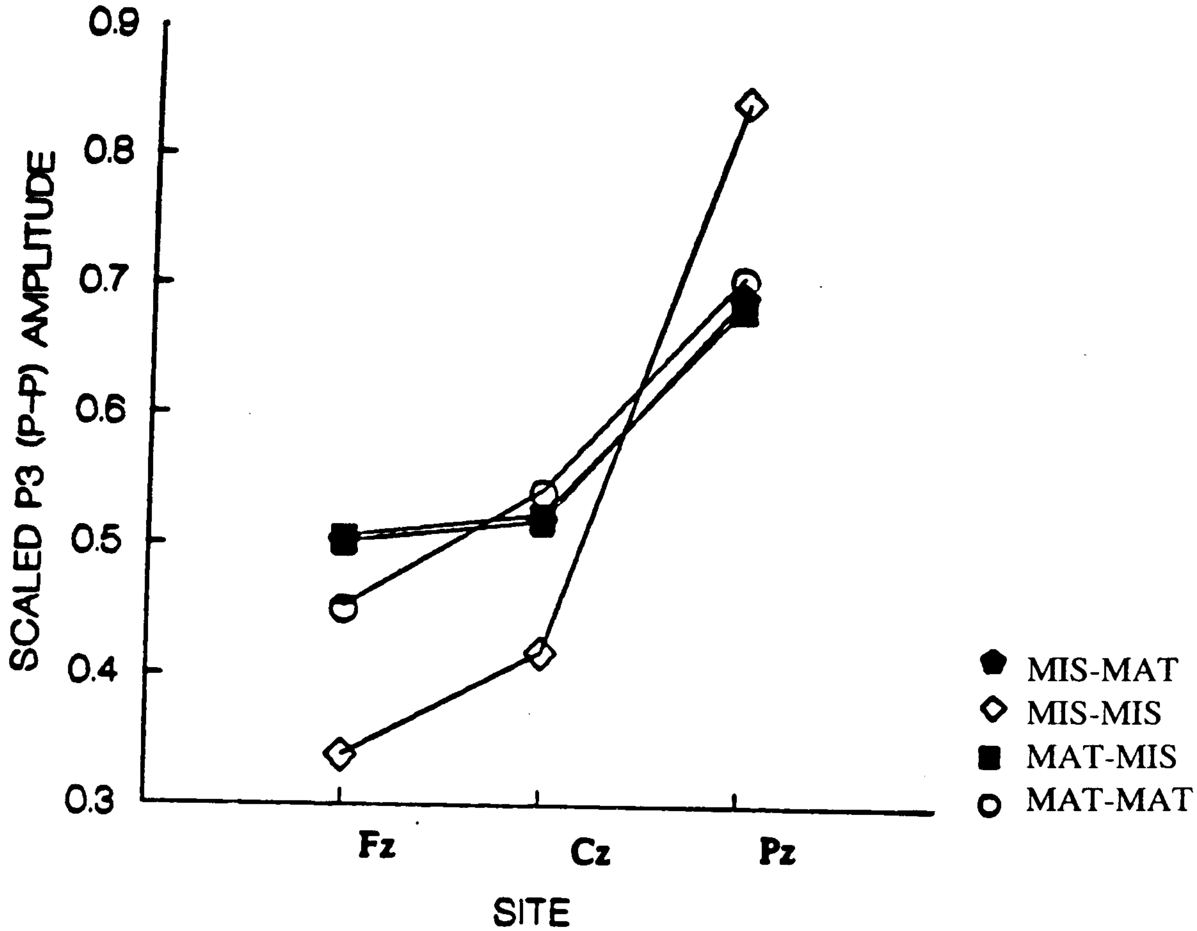

Fig. 5. Scaled P3 group mean amplitudes at three sites from two experiments, the "Birthday" (autobiographical) study and the "FCP" (match-to-sample or forced choice procedure) study for liar ("malingerer") and truth-telling groups in both studies. Only responses to oddballs and matches are shown.

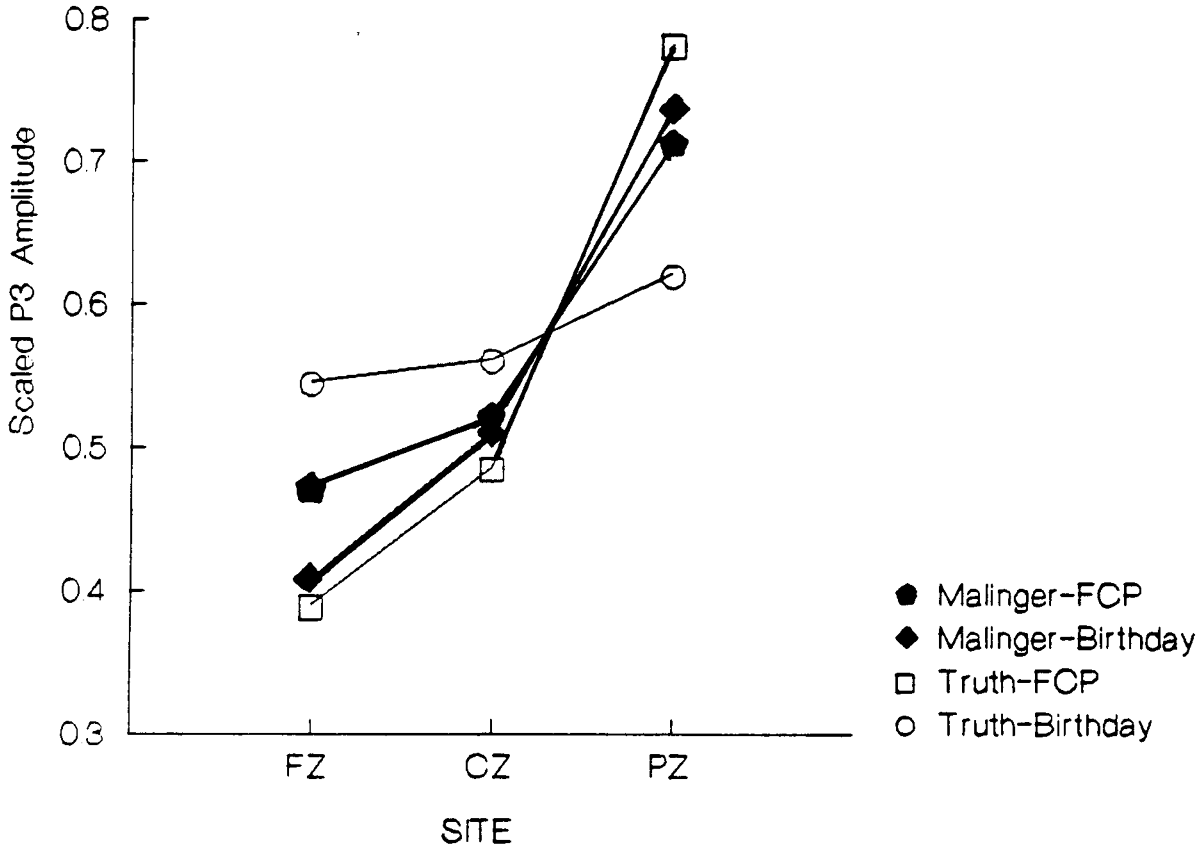

Fig. 6. (From Miller, 1999a.) Scaled group mean P3 amplitude as a function of site in two trial blocks; the first ("Phone") where both liars (called "malinger") and truth-telling subjects told the truth, and the second ("Birthday") where liars lied and truth-tellers continued to tell the truth. Only responses to autobiographical oddballs are shown.

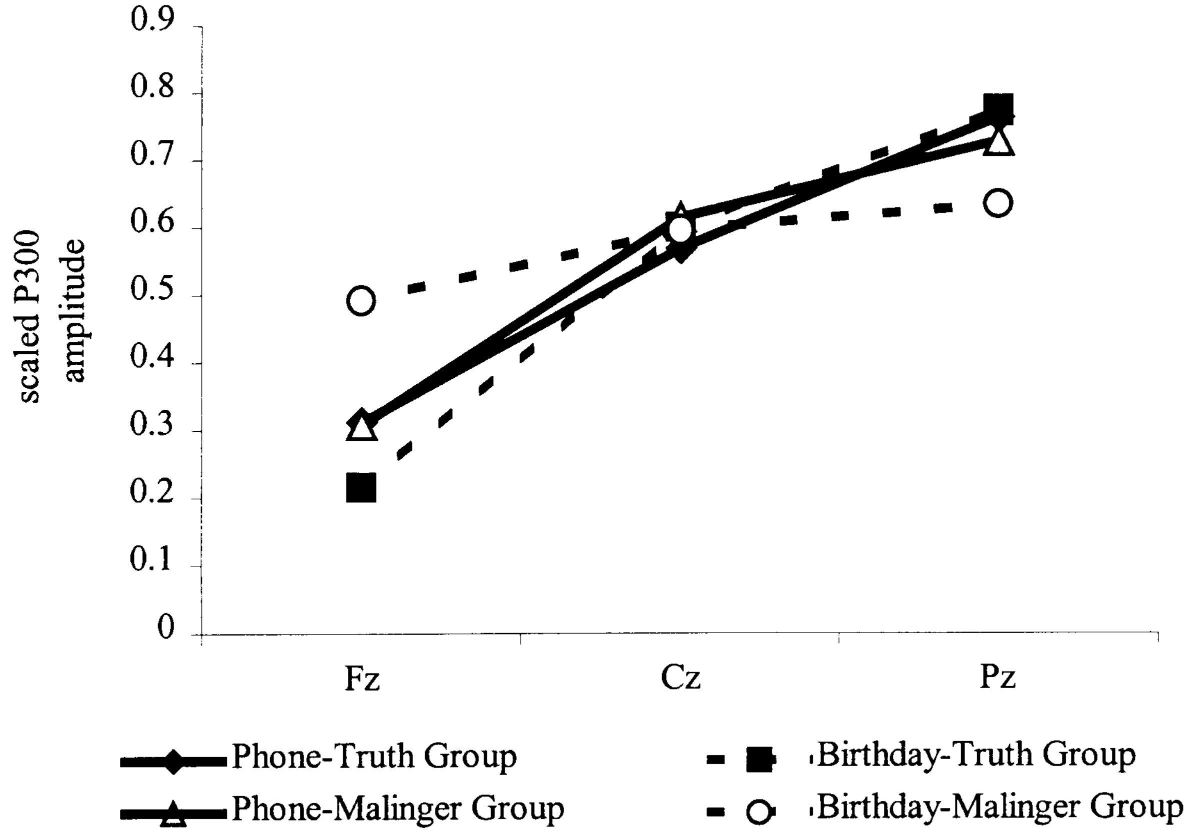

Fig. 7. (From Miller, 1999a.) Scaled mean P3 amplitudes in the liar group, second ("Birthday") block on trials in which they told the truth ("yes") and on trials in which they lied ("no"). Only responses to oddballs are shown.

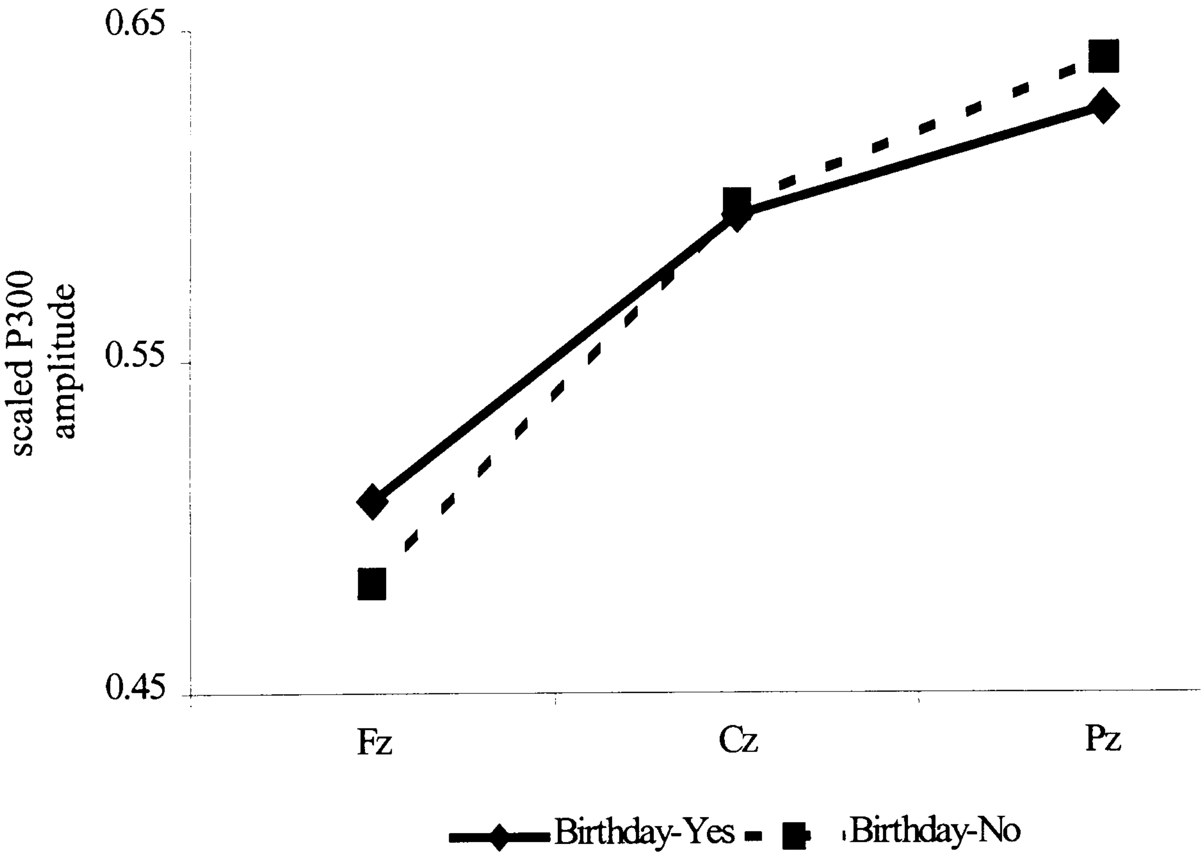

Fig. 8. (From Miller, 1999b.) Scaled group mean P3 profiles for first block (“Phone”) and second block (“Birthday”) in response to autobiographical oddball stimuli in “Low” and “High” psychopathy groups, both lying, as a function of scalp site. No interactions with site are apparent.

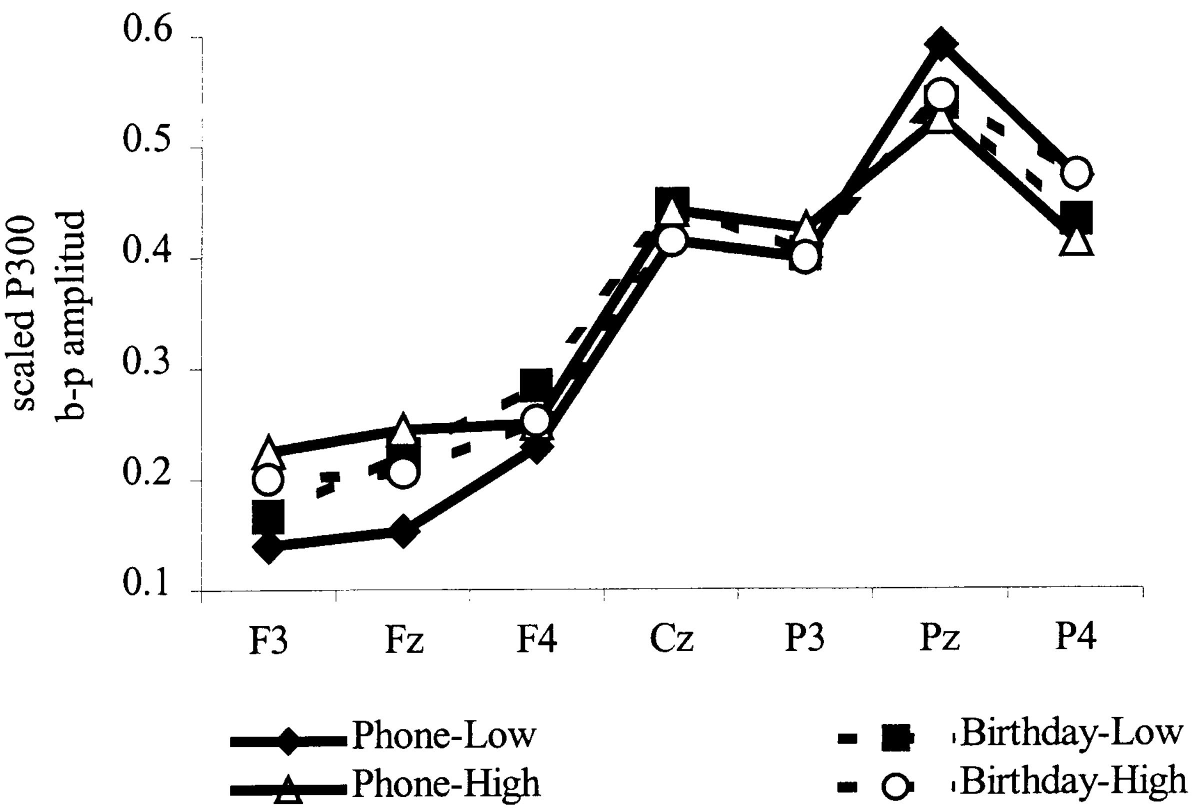

Fig. 9. (From Miller, 1999b.) Data as in Fig. 7, Birthday block only, but sorted by response type (honest = “yes” and dishonest = “no”). The dishonest and honest response curves are not parallel for both high and low psychopathy groups.

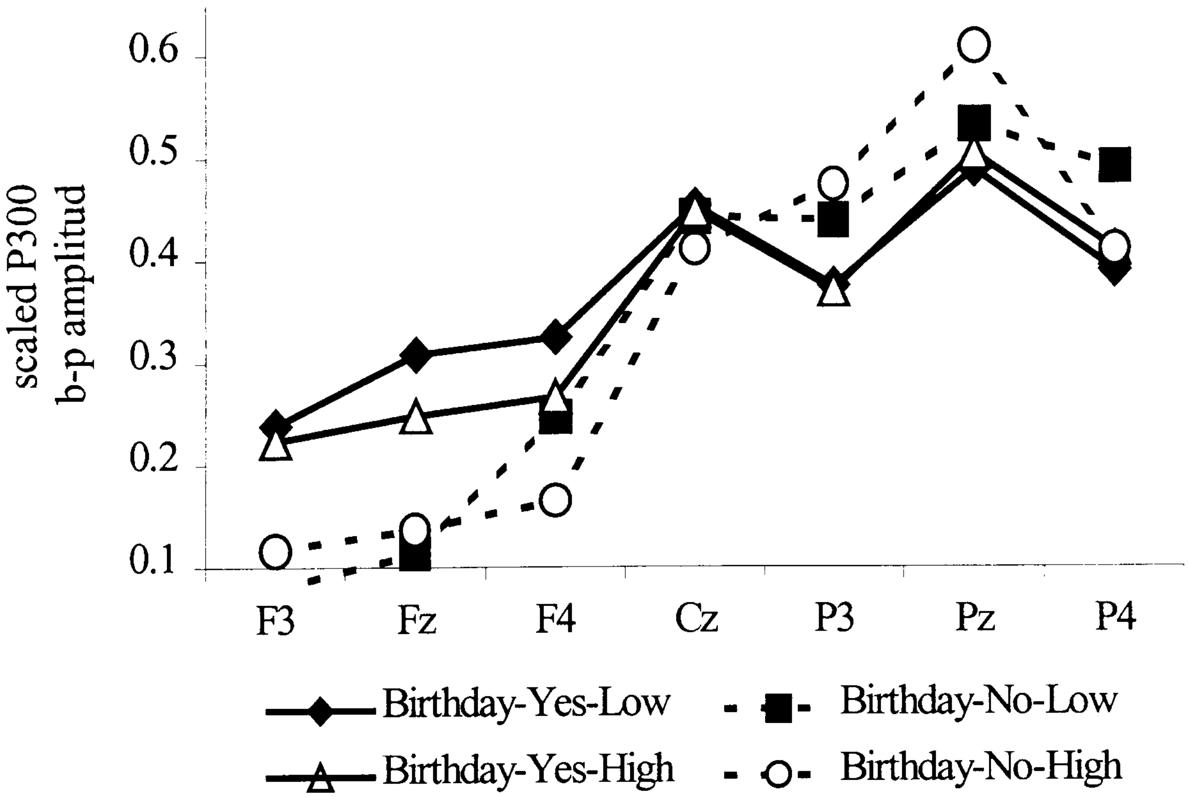

Fig. 10. Scaled P3 amplitude as a function of site in control (B1, B2) and liar groups (L1, L2) in response to autobiographical oddball stimuli. The numbers indicate first (B1, L1) or second (B2, L2) blocks. The filled symbols are for dishonest responses in the liar group (L2) or honest responses repeated backwards in the control group (B2). Unfilled symbols are for both groups on honest, forwards-repeating trials. It is seen that only liars, when they lie, produce an aberant profile (filled diamonds).

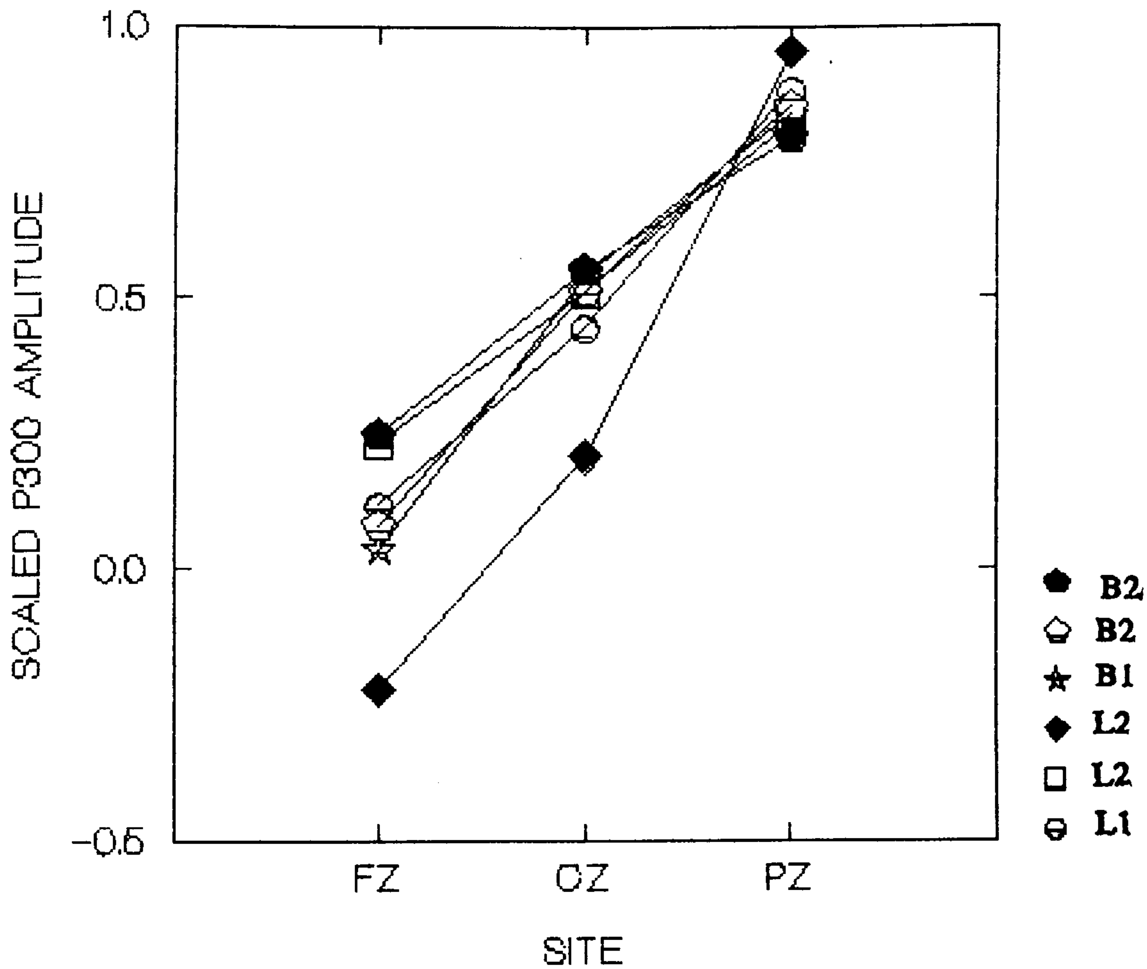

Fig. 11. (From Miller, 1999c.) Mean scaled P300 amplitudes in response to Old, Lure-Old, Lure-New, and New words.

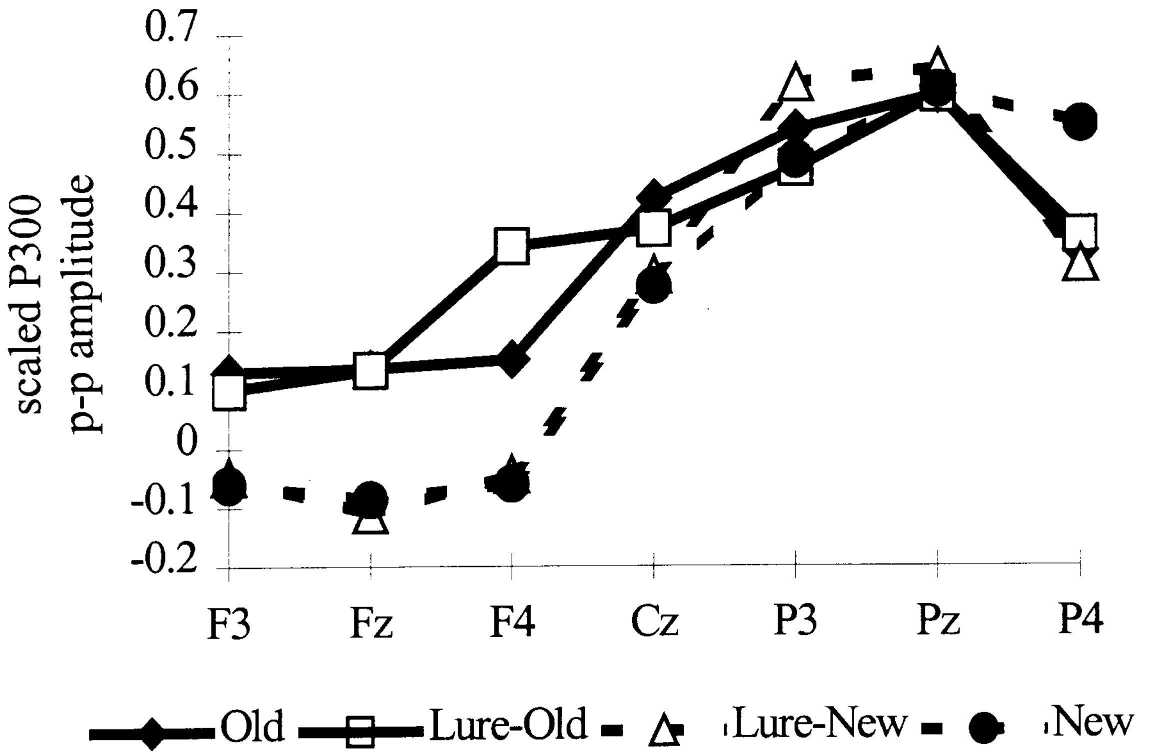

Fig. 12. (From Miller 1999c.) Mean P300 latencies at site Pz in response to words as in Fig. 11.