

Event-related brain potentials and subjective probability in a learning task

RICHARD L. HORST, RAY JOHNSON, JR., and EMANUEL DONCHIN
Cognitive Psychophysiology Laboratory, University of Illinois, Champaign, Illinois 61820

In a paired associate learning task, subjects responded to each presentation of a nonsense syllable by typing both a three-letter associate and a rating of their confidence that this response was correct or incorrect. Average event-related potentials (ERPs) elicited by the subsequent presentation of the actual paired syllable varied with the interaction of confidence and trial outcome. A larger amplitude P300 was elicited by syllables that informed subjects that they were correct when they thought they were incorrect or that they were incorrect when they thought they were correct than by syllables that confirmed subjects' expectations. That this average ERP result was indeed an effect on P300 amplitude, and not an artifact of single-trial variability in P300 latency, was confirmed with a trial-by-trial latency adjustment procedure. Consistent with findings from other tasks, $\overline{P300}$ amplitude varied inversely with the subjective probability of the ERP-eliciting events.

An event-related brain potential (ERP) is a transient sequence of voltage fluctuations induced in brain tissue by the occurrence of some critical event. In scalp recordings, the ERPs are often obscured by the substantially larger amplitude of the ongoing electrical activity of the brain. However, reliable estimates of the waveform of the ERP can be obtained by applying signal-averaging techniques to electroencephalographic data (see Callaway, Tueting, & Koslow, 1978). The waveform of the ERP, that is, the sequence and relative amplitude of the negative and positive peaks, depends on the eliciting events, on the way these events are processed by the subject, and on the placement of the recording electrodes on the scalp. Often, deflections in the ERP that appear at consistent temporal intervals following the eliciting events are called components. Components are labeled by a <letter> <number> combination—the letter indicating the polarity of the component (positive or negative) and the number representing its minimal, or modal, latency (see Donchin, Ritter, & McCallum, 1978, for a discussion of components).

The work described here was supported, in part, by DARPA, through Contract N00014-76-C-0002 with the Office of Naval Research. A preliminary report of this study was presented at the meeting of the Society for Psychophysiological Research, Madison, Wisconsin, 1978. We are grateful to Paul Lamprinos, who did much of the PLATO programming and implemented the PLATO-PDP-11/10 interface, to Marta Kutas and Lisa Brenner, who were involved in pilot work for the present study, and to Theodore Bashore, John Polich, and Christopher Wickens for their helpful comments on the manuscript. Richard L. Horst's present address is Applied Neuroscience Institute, University of Maryland Eastern Shore, Princess Anne, Maryland 21853. Ray Johnson's present address is Stanford University School of Medicine, Department of Psychiatry and Behavioral Sciences, Stanford, California 94305. Address reprint requests to Emanuel Donchin.

Many components are "exogenous." That is, they are scalp manifestations of the obligatory responses of brain tissue to sensory input. The pattern of exogenous components is strongly dependent on the physical properties of the eliciting stimuli and on the state of the subject's sensory systems. In some circumstances, the exogenous components are followed by "endogenous" components. These endogenous components are manifestations of information processing activities and are therefore determined by the psychological context in which the eliciting events occur, rather than by the physical nature of the stimuli. The family of endogenous components thus provides a set of novel tools that may be employed in the study of cognitive function (Donchin, 1979; Donchin & Isreal, in press; Hillyard, Picton, & Regan, 1978).

The P300, as one of these components is called, appears to be a manifestation at the scalp of brain activity invoked by the occurrence of task-relevant events (Donchin & Cohen, 1967; Sutton, Tueting, Zubin, & John, 1967) that resolve, for the subject, some uncertainty (Sutton, Braren, Zubin, & John, 1965). Furthermore, the latency of P300 is proportional to the time it takes the subject to categorize the eliciting event (Kutas, McCarthy, & Donchin, 1977; Ritter, Simson, & Vaughan, 1972; N. Squires, Donchin, K. Squires, & Grossberg, 1977).

Much evidence supports the assertion that the amplitude of $\overline{P300}$ varies inversely with the probability that the subject associates with the eliciting event. A completely predictable event, even if task relevant, elicits little if any $\overline{P300}$ (Donchin, Kubovy, Kutas, Johnson, & Herning, 1973; Friedman, Hakerem, Sutton, & Fleiss, 1973; Sutton et al., 1965). When there is uncertainty as to which of two events will occur, the less frequently occurring event elicits the larger $\overline{P300}$ (Sutton et al.,

1965). Moreover, when event probability is manipulated, systematic variations in $\overline{P300}$ amplitude are obtained. Tueting, Sutton, and Zubin (1970), using a guessing task, were the first to show that as the prior probability of a stimulus was decreased, the amplitude of the elicited $\overline{P300}$ increased (also see, e.g., Friedman et al., 1973; K. Squires, Donchin, Herning, & McCarthy, 1977). By parametrically varying stimulus probabilities in a counting task, Duncan-Johnson and Donchin (1977) demonstrated that $\overline{P300}$ amplitude, for task-relevant stimuli, was a decreasing function of prior probability over a range from .10 to .90.

In addition to the effect of prior probability, $\overline{P300}$ amplitude varies with the sequence of preceding events (see Tueting et al., 1970). At all levels of probability in the Duncan-Johnson and Donchin (1977) study, a stimulus that had been preceded by itself elicited a smaller $\overline{P300}$ than one that had been preceded by the other stimulus. Similarly, K. Squires, Petuchowski, Wickens, and Donchin (1977) and K. Squires, Wickens, N. Squires, and Donchin (1976) showed that the $\overline{P300}$ elicited by a stimulus in a Bernoulli series is influenced by the sequence of stimuli presented on the preceding five trials. K. Squires et al. (1976) suggested a single construct to account for both these effects on $\overline{P300}$. They proposed that the subjective probability (or "expectancy") associated with an event is a linear combination of the prior probability of that event and the subject's exponentially decaying memory of the sequence of preceding events. Assuming that $\overline{P300}$ amplitude is inversely related to this subjective probability, their model accounted for 78% of the variance in $\overline{P300}$ amplitude. Similar models have been developed to account for sequential effects in choice reaction time (RT) (Audley, 1973; Falmagne, 1965; Laming, 1969).

The effects of event probability and sequence on $\overline{P300}$ cannot be attributed to habituation or to receptor adaptation. It appears to be the probability of stimulus categories, rather than the frequency with which particular physical stimuli occur, that governs the effects of both prior probability (Courchesne, Hillyard, & Courchesne, 1977; Friedman, Simson, Ritter, & Rapin, 1975; Kutas & Donchin, 1978; Tueting et al., 1970) and event sequence (Johnson & Donchin, in press) on $\overline{P300}$. The term "category" refers here to classifications that are imposed on the stimuli by the subject due to the structure of the task. Furthermore, the degree to which the previous sequence of stimuli affects the amplitude of $\overline{P300}$ depends on task conditions. In a warned RT task, Duncan-Johnson and Donchin (1978) showed that sequential effects on $\overline{P300}$ were eliminated when the warning stimulus provided information about the probability with which particular imperative stimuli would occur.

In almost all previous studies in which the relation between subjective probability and $\overline{P300}$ amplitude was examined, subjects derived the expectancies that they presumably assigned to events from attributes of the

environment—the prior probabilities and the sequences in which the experimenter delivered stimuli. In the present study, we attempted to determine if the relationship between $\overline{P300}$ amplitude and subjective probability would hold when subjects formed expectancies on the basis of their changing knowledge about the environment. The subjects were assigned a classical paired associate learning task. In response to the first ("stimulus") syllable of each pair, subjects typed the three-letter syllable that they thought was the paired associate. They also reported their confidence in the correctness of this response. The correct paired ("response") syllable was then presented. The extent to which the correct "response" syllable was expected at this point was assumed to depend on the subjects' confidence in the correctness of their three-letter responses.¹ As learning occurred, these internally formed expectancies should have changed, even though there was no change in the manner with which external stimuli were being presented. Thus an analysis of the ERPs elicited by the "response" syllables according to the subjects' confidence ratings and to the trial outcomes (that is, whether or not their three-letter responses had in fact been correct) allowed an examination of the relationship between subjective probability, as inferred from subjects' own indications of their expectancies, and $\overline{P300}$.

K. Squires, Hillyard, and Lindsay (1973) have studied the amplitude of $\overline{P300}$ elicited by stimuli that indicated to subjects whether they had been correct or incorrect in the detection of a near-threshold auditory stimulus. They report that the amplitude of $\overline{P300}$ was larger when these feedback stimuli disconfirmed subjects' judgments. Whether the same effect would be obtained in a learning task in which the ERP-eliciting stimuli provided feedback as to the accuracy of associations being formed in memory, rather than as to the accuracy of a sensory discrimination, was of interest here.

METHOD

Subjects

Six students at the University of Illinois (three males) were paid for their participation in the experiment. Their ages ranged from 19 to 28 years. Four subjects had participated in previous ERP experiments. A seventh subject completed all three sessions, but his data were discarded because his confidence ratings were confined almost exclusively to the two extreme points of the rating scale.

Apparatus and Stimuli

Subjects sat in an easy chair positioned in front of a PLATO computer terminal (see Smith & Sherwood, 1976) and held a detachable keyboard in their laps. The ERP-eliciting stimuli were consonant-vowel-consonant (CVC) nonsense syllables presented on the plasma-panel display of the terminal (see Johnson, Bitzer, & Slottow, 1971). The CVCs subtended .6 x .2 deg of visual angle and were 3.2 fL in luminance, compared with the .2-fL background of the display. A continuously presented rectangle, which subtended 2.8 x 1.2 deg of visual angle, surrounded the area of the panel at which the CVCs appeared and served as a target for the subject's gaze. Ambient

lighting was adjusted to a comfortable level for each subject.

The subjects learned from repeated presentations which "response" CVC was paired with each "stimulus" CVC. Lists of six paired CVCs were constructed with the following constraints: (1) All CVCs were of low meaningfulness (less than or equal to 1.50 on the m' scale; see Noble, 1961); (2) no CVC appeared in more than one list; (3) the six "stimulus" CVCs were highly similar, usually differing from each other in only one or two letters; (4) the six "response" CVCs were much less similar—no two of them had the same consonant in a given position and no syllable contained any letters of the paired "stimulus" CVC.

The PLATO computer system controlled the presentation of stimuli and processed subjects' responses from the keyboard. A PDP-11/10 received synchronizing pulses and identifying information from the PLATO computer, digitized and processed the EEG, and allowed the experimenter to monitor data collection via a GT-40 display. Data analyses were performed off-line on a Harris /7 computer. The statistical packages SPSS (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) and ALICE (Grubin, Bauer, & Walker, 1976) were used for data analysis.

Procedure for Paired Associate Task

Events on each trial. As illustrated in Figure 1, after a 1,000-msec foreperiod, during which the target rectangle was empty, a "stimulus" CVC was presented for 500 msec. Then, following a 1,000-msec delay, three question marks were displayed in the rectangle, signaling the subject to respond. The subject then typed the three letters that he or she thought was the correct "response" CVC, followed by a confidence rating from 0 to 100. The subjects' responses were echoed on the PLATO display and appeared in the rectangle. The key-stroke that terminated the confidence rating initiated a 1,000-msec interval, during which the rectangle was again empty. The correct "response" CVC was then presented for 500 msec. After a further delay of 1,000 msec, three percent signs appeared in the rectangle, signaling a 4-sec intertrial interval (ITI). The offset of these percent signs initiated the next trial. If the subject struck any key before the three question marks appeared, failed to complete the responses within 15 sec, or entered an invalid confidence rating, three asterisks were displayed instead of the "response" CVC. ERPs were recorded both to presentations of the "stimulus" and "response" CVCs. In each case, the recording epoch extended for 1,750 msec, starting 250 msec before CVC onset.

Learning paired associate lists. The pair of syllables to be presented on each trial was selected at random from the five pairs in the list, excluding the one that had been presented on the previous trial. This procedure was followed until the subject gave two consecutive correct responses to each of the six pairs in the list. If a subject, after twice responding correctly to a given "stimulus" CVC, subsequently responded to it incorrectly, two further correct responses were required. All subjects learned

the same eight lists, two in the first session and three in both the second and third sessions, but in a randomized order.

Instructions. Before starting to learn each list, subjects were reminded to watch the target rectangle and, from the beginning of each CVC foreperiod until the question marks or percent signs appeared following a CVC, to avoid movements of the eyes, mouth, or body that could cause recording artifacts. The following instructions regarding the use of the confidence rating scale also appeared on the PLATO terminal prior to the presentation of each list: "We want to correlate your brain waves with your confidence ratings. So it is very important that on every trial you do the confidence rating as accurately as you can. Remember after entering a three-letter response you are to rate your confidence as to whether that response, as a whole, was correct or incorrect. The confidence scale is meant to represent a continuum of confidence from one extreme, where you are as sure as you can be that your response was incorrect (0—definitely incorrect), to the other extreme, where you are as sure as you can be that your response was correct (100—definitely correct).

"As a general guideline, use a rating between 0 and 25 when you are very sure that your three-letter response was incorrect; use a rating between 25 and 50 when you think your response was probably incorrect, but you are not so sure; use a rating between 50 and 75 when you think your response was probably correct, but you are not sure; use a rating between 75 and 100 when you are very sure that your response was correct.

"Within these general guidelines, you should choose an integer which you feel reflects your confidence accurately, with relatively large numbers indicating more likely correct and relatively small numbers indicating more likely incorrect.

"Remember that you should try to learn each list as fast as possible. If you have any questions, ask the experimenter now."

Determining confidence ranges. Pilot work, in which a 4-point confidence rating was used, revealed marked individual differences in the manner with which subjects rate their confidence in the paired associate task. Since the same numerical value appears to have different meanings to different subjects, it would be misleading to use the nominal values of the confidence ratings to classify the ERPs. In this study, we used a 101-point confidence scale. This choice allowed us to partition each subject's scale, based on that subject's usage of the scale, into ranges that would be equivalent across subjects.

With the following procedure, each subject's data were partitioned into four such ranges of confidence. First, the 101-point scale was collapsed to 21 points by combining the ratings in successive 5-point sections of the scale (Rating 100 was treated as a "section" by itself). The ratings in these sections were then further grouped into "regions" of the scale (Figure 2a) that each contained 4% or more of all the ratings entered by that subject while learning all eight lists.² Next we determined the percentage of trials in each of these regions on which the subject entered the correct three-letter response (Figure 2b). Finally, with the constraint that only adjacent regions could be combined, the scale was further collapsed into four "ranges" of confidence such that the combined trials best approximated 0%, 33%, 67%, and 100% correct (Figure 2c).

This partitioning resulted in ranges of confidence that can be considered equivalent, in terms of percentage of correct trials, across subjects. Note that the partitioning was done only as a matter of convenience for examining averaged ERPs. No claim is made that the derived ranges correspond, in either number or boundaries, to confidence ranges that the subjects may have formed internally. Note further that since the partitioning was done without regard to the ERP data, we did not prejudice the existence of ERP differences among the four confidence ranges. For convenience, we will refer to the four ranges of confidence, those at which accuracy approximated 0%, 33%, 67%, and 100%, as, respectively, the "certainly wrong," "probably wrong," "probably right," and "certainly right" ranges; however, we imply neither that the trials within a given range are homo-

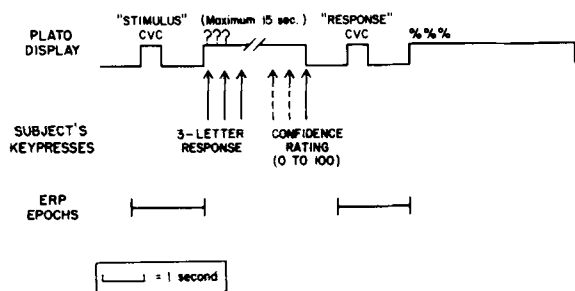


Figure 1. The events within each trial of the paired associate learning task.

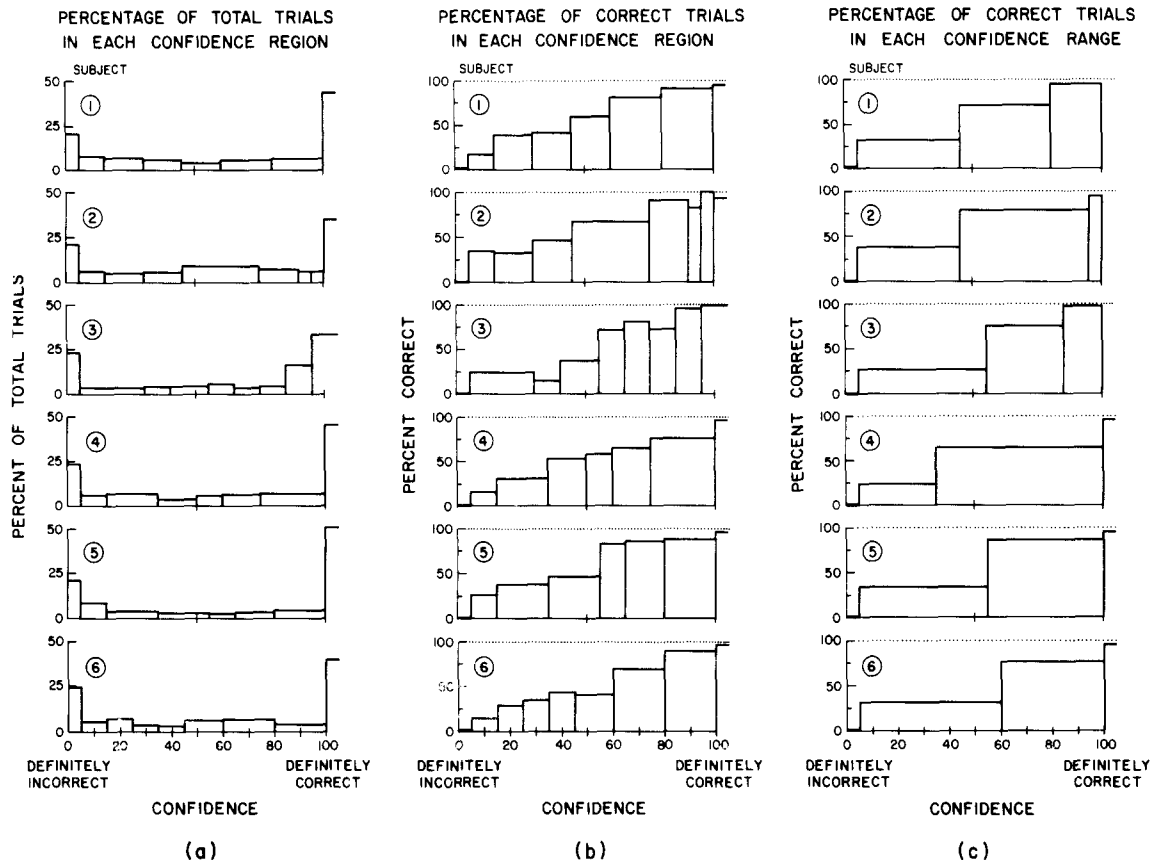


Figure 2. (a) The percentage of trials on which each subject used each region of the confidence scale. These regions are groups of ratings that together contain at least 4% of the total trials for a given subject. The bar extending to the right of some graphs indicates that Rating 100 was itself a "region" for that subject. (b) For each region and subject, the percentage of the trials on which the correct three-letter response was made. (c) For each subject, the partitioning of the confidence scale that resulted from collapsing regions into the four ranges of confidence that best approximated 0%, 33%, 67%, and 100% correct. The bar extending to the right of some graphs indicates that Rating 100 was itself a "range" for that subject.

geneous nor that the ranges necessarily represent symmetrical states of confidence.

Procedure for Counting Task

For comparison with the ERPs recorded in the paired associate task, we obtained ERPs while the subjects counted the number of times a specific CVC occurred in a random sequence of six CVCs. This task was included because a well-defined P300 is typically seen in such a task. Lists of six single CVCs were constructed with the same constraints as the "response" CVCs of the paired associate lists. No CVC appeared in both the count and paired associate lists.

Each trial consisted of a 1,000-msec foreperiod, followed by the 500-msec presentation of a randomly selected CVC (other than the one that had just occurred). Then, following a 1,000-msec delay, three question marks appeared in the target rectangle, signaling a 4-sec ITI. With the disappearance of the question marks, the foreperiod of the next CVC began. As in the paired associate task, ERPs and eye movements were recorded for 1,750 msec, beginning 250 msec before CVC onset. A block of 60 counting task trials was presented at the beginning and end of each experimental session. Prior to each block, one of the six CVCs in the list was designated as the target, and subjects were asked to keep a covert count of the number of times it occurred. At the end of the block, subjects typed their counts (these were always accurate to within plus or minus one).

Recording

EEG was recorded from frontal, central, parietal, and occipital scalp sites (Fz, Cz, Pz, and Oz in the International 10-20 system), each referred to the linked mastoids. The electro-oculogram (EOG) was recorded from suborbital and supraorbital sites, each referred to the linked mastoids. Subjects were grounded with a chin electrode. Burden Ag-AgCl electrodes, affixed with collodion, were used on the scalp. Beckman Bio-potential electrodes affixed with adhesive collars were used for the EOG, ground, and reference sites. Electrode impedances were always below 10 kohms. EEG and EOG were amplified by modified Grass Model 7P122 amplifiers (with an upper half-amplitude of 35 Hz and a time constant of 8 sec). The PDP-11/10 sampled the EEG and EOG every 10 msec during the 1,750-msec epochs. These digitized ERPs, along with identifying information, were written on magnetic tape.

Analysis of ERPs

Trials with EOG activity sufficient to contaminate the scalp recordings were identified with a peak-detection algorithm. Only trials free of contamination were included in the ERP analyses, whereas all trials were included in the analyses of behavioral data. Since variability in the latency of P300 among the paired associate average ERPs made a principal-component analysis of the waveforms inappropriate (see Donchin & Heffley, 1978), a base-to-peak amplitude measure of P300 was employed. Since

it was necessary to compare average ERPs that were composed of very different numbers of trials, average ERPs were first digitally low-pass filtered (half-power frequency = 6.3 Hz; see Ruchkin & Glaser, 1978) to attenuate any high-frequency EEG activity that remained in the averages. Then the difference between the mean voltage of the prestimulus ERP points and the voltage of the most positive point between 350 and 950 msec after CVC onset was calculated.

RESULTS

Paired Associate Behavioral Data

Trials to criterion. There was considerable variability both within and between subjects in the number of trials needed to learn a list. Across subjects, the mean number of trials to criterion was 56 (SD = 21). Repeated-measures analyses of variance showed no systematic differences either within or across sessions in the number of trials to criterion.

Confidence ranges and stages of learning. Since we wish to infer subjects' expectancies for "response" CVCs from their confidence ratings, it is necessary to provide evidence that the confidence ratings were valid. If the ratings actually did reflect subjects' knowledge about the paired associates, relatively high numerical ratings should have been concurrent with relatively accurate three-letter responses. Figure 2b shows that, in general, the percentage of correct responses increased with numerically increasing confidence ratings for each subject. Furthermore, the incidence of ratings in the four confidence ranges should have changed as learning progressed. As subjects changed from consistently responding incorrectly to consistently responding correctly to a given "stimulus" CVC, their confidence should have shifted systematically along the scale from numerically low to numerically high ratings. To investigate this possibility, we divided all presentations of each CVC pair to each subject into three "stages" of learning: (1) trials prior to the first correct response for the pair, (2) trials from the first correct response until the last incorrect response, and (3) trials following the last incorrect response (pairs that were always responded to correctly after the first correct response contributed no trials to Stage 2). Table 1 shows, at each stage of learning, the percentage of ratings in each of the four confidence ranges, averaged over subjects and CVC pairs. Before responding correctly to a given CVC, subjects tended to indicate that they were wrong; when consistently responding correctly, they tended to indicate

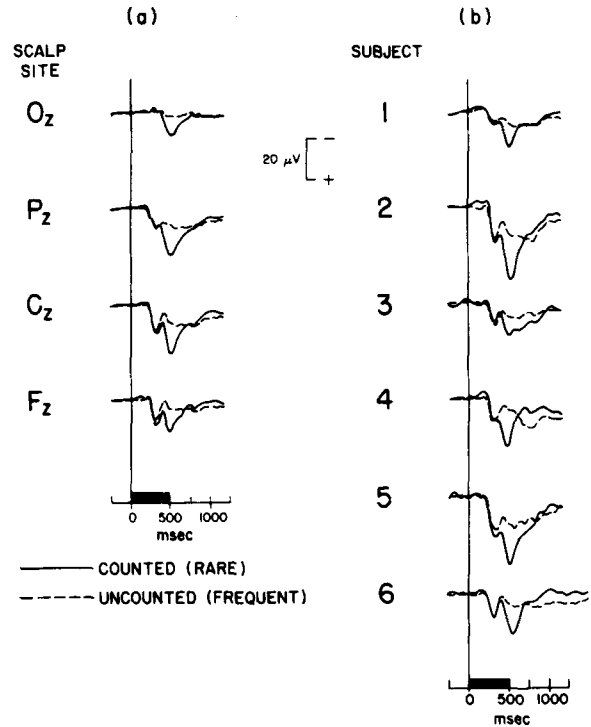


Figure 3. (a) Grand-averaged (over subjects) ERPs from the counting task. At each scalp site, the ERPs elicited by counted and uncounted CVCs are superimposed. (b) Digitally filtered average ERPs from Cz for each subject. ERPs elicited by the counted and uncounted CVCs are superimposed.

that they were right; when responding to a CVC pair with inconsistent accuracy, their ratings were more evenly distributed.

Thus subjects' confidence ratings appear to be a valid index of their knowledge. It is reasonable to assume, therefore, that when subjects indicated that they were "probably right" or "certainly right," they would have expected the "response" CVC to inform them that their three-letter response was correct; conversely, when subjects indicated that they were "probably wrong" or "certainly wrong," they would have expected the "response" CVC to inform them that their three-letter response was incorrect.

Average ERPs

Counting task. In Figure 3a the ERPs that were elicited by counted and uncounted CVCs are superimposed. These ERPs have been grand-averaged over sub-

Table 1
At Each Stage of Learning: The Percentage of Trials in Each Confidence Range (Averaged Over Subjects)

Stage of Learning	Confidence Range			
	CW	PW	PR	CR
Before first time correct	62.5	28.9	7.3	1.3
From first time correct to last time incorrect	15.8	28.5	25.4	30.4
After last time incorrect	.2	5.3	14.8	79.8

Note—CW = certainly wrong; PW = probably wrong; PR = probably right; CR = certainly right.

jects and blocks of trials. Two positive-going waves with different scalp distributions are prominent. One (P280) is larger at the central and frontal sites and appears equally in the ERPs elicited by the counted and uncounted CVCs. The later positivity (400 to 700 msec after CVC onset) has a centroparietal maximum and is apparent only in the ERPs elicited by the counted CVCs. This difference in late positivity was observed in each subject's ERPs (Figure 3b). Since the probability of the counted CVC was 16.7% and that of the uncounted CVCs combined was 83.3%, this late positivity seems to be the centroparietal P300³ that is elicited by task-relevant rare events (see review by Donchin et al., 1978).

As is typically the case, there were individual differences in the scalp distribution of P300. For comparison with the ERPs from the paired associate task, these scalp distributions were expressed as percentages of maximum base-to-peak amplitude and are presented in Table 2.

Paired associate task. For each subject, the ERPs elicited by the "stimulus" and the "response" CVCs were each averaged separately for eight categories of trials (ratings in each of the four confidence ranges by two trial outcomes). By necessity, the number of trials in these various categories differed markedly (see Figure 2). Since confidence judgments in the "certainly wrong" range were rarely used with correct three-letter

Table 2
Scalp Distribution of P300 in the Two Tasks for Each Subject (Percent of Maximum Base-to-Peak Amplitude)

Subject	Scalp Site							
	Counting Task				Paired Associate Task			
	Fz	Cz	Pz	Oz	Fz	Cz	Pz	Oz
1	46	89	100	65	52	96	100	59
2	63	100	90	48	71	100	84	37
3	65	99	100	63	62	100	98	52
4	48	94	100	56	52	95	100	56
5	73	100	83	24	78	100	78	26
6	64	100	90	23	64	100	91	24

Note—Each scalp measure for the counting task is based on P300 amplitude in the average ERPs elicited by the counted (rare) CVC. Each scalp measure for the paired associate task is based on the grand mean P300 amplitude computed over the seven average ERPs elicited by the "response" CVCs in the various confidence range by trial outcome categories.

responses, we could not compute reliable average ERPs for the "certainly wrong"/correct category.

The ERPs elicited by "stimulus" and "response" CVCs in the seven remaining categories (grand-averaged over subjects at each of the four scalp sites) are superimposed in Figure 4. A P280 wave, similar to that appearing in the counting task ERPs, is seen in both the "stimulus" and "response" CVC ERPs. There were no consistent differences in either the latency or amplitude

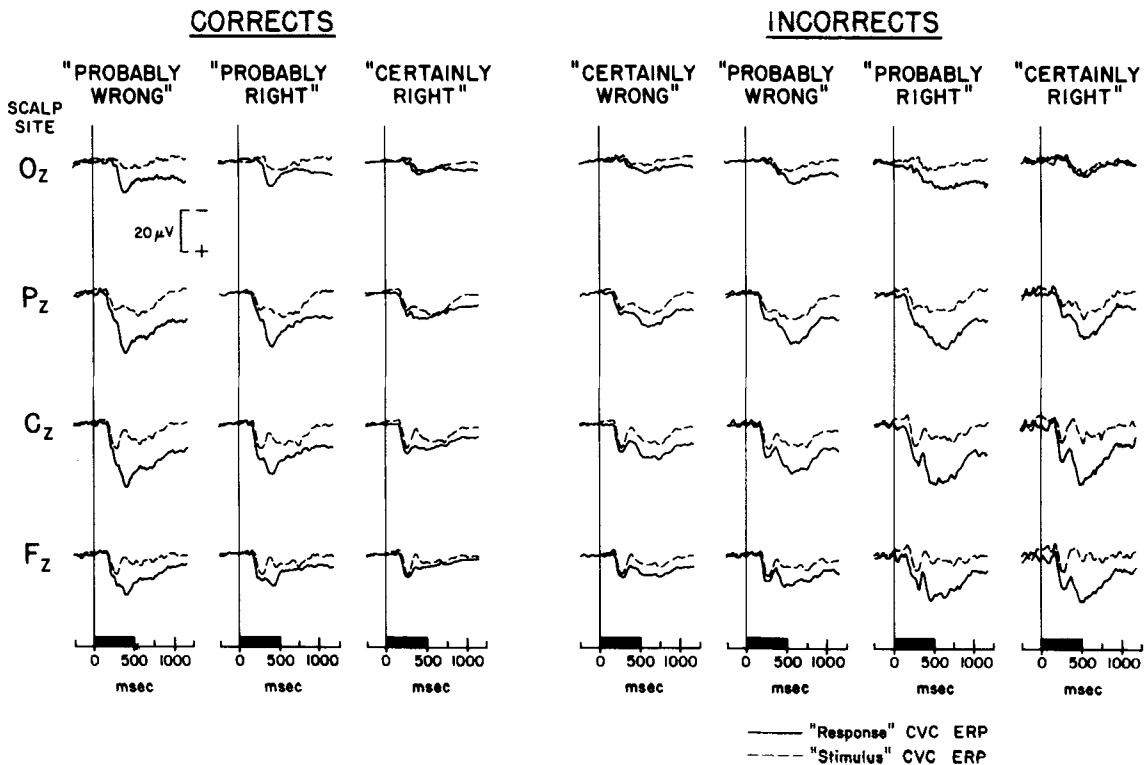


Figure 4. Grand-averaged (over subjects) ERPs elicited by the "stimulus" and "response" CVCs. Separate averages are shown for trials on which subjects rated their confidence in each of the four ranges and when their three-letter responses were correct and incorrect. There was an insufficient number of trials in the "certainly wrong"/correct category to consider.

of this wave among the seven categories. The "stimulus" CVC ERPs displayed relatively little late positivity and, in contrast to the report of Peters, Billinger, and Knott (1977), did not vary systematically in base-to-peak amplitude among the seven categories.

In the "response" CVC waveforms, however, a substantial late positivity with a central-parietal maximum was apparent. Moreover, there was considerable variability in both the amplitude and peak latency of this late positivity among the categories. On correct trials the positivity was largest when the rating was in the "probably wrong" range, and it decreased with increasing confidence that the three-letter response was correct. On incorrect trials the positivity was larger for "certainly right" and "probably right" ratings, and it decreased with increasing confidence that the three-letter response was incorrect. These trends were pronounced to the extent that at the "probably wrong" confidence level, a larger amplitude late positivity was elicited by the "response" CVC on correct trials than on incorrect trials, whereas at both "probably right" and "certainly right" levels of confidence, a larger late positivity was elicited on incorrect trials than on correct trials. That these trends were consistent across subjects is shown in Figure 5, in which the ERPs elicited by correct and incorrect "response" CVCs are superimposed for each subject at the different confidence levels.

Particularly striking in Figures 4 and 5 is the breadth and sometimes multi-peaked form of the late positivity. It is possible that these average ERPs reflect a sharper-peaked P300 (such as that seen in the counting task) that varied considerably in latency from trial to trial. But it is also possible that the late positivity in the paired associate ERPs was composed of multiple positive ERP components (see Friedman, Vaughan, & Erlenmeyer-Kimling, 1978; Goodin, K. Squires, Henderson, & Starr, 1978; Roth, Ford, & Kopell, 1978; Stuss & Picton, 1978). Inspection of individual subject's average waveforms across scalp sites failed to reveal any consistent differences in the scalp distribution either of the various peaks in the late positivity or of the peak positivity among the seven categories. Moreover, individual differences in the scalp distribution of the late positivity in the "response" CVC ERPs (Table 2) conformed remarkably to those seen in the counting task (the base-to-peak amplitudes at the four scalp sites had a product-moment correlation between tasks of .87). Thus we found no indication that the broad late positivities in the paired associate ERPs reflect anything other than a P300 that varied in latency from trial to trial.

Mean base-to-peak amplitudes of these P300s in the average ERPs from Cz are presented in Table 3. Repeated-measures analysis of variance of these base-to-peak amplitudes (six subjects with repeated measures on two trial outcomes \times three confidence ranges; the "certainly wrong"/incorrect category was excluded) indicated that the Confidence Range by Trial Outcome

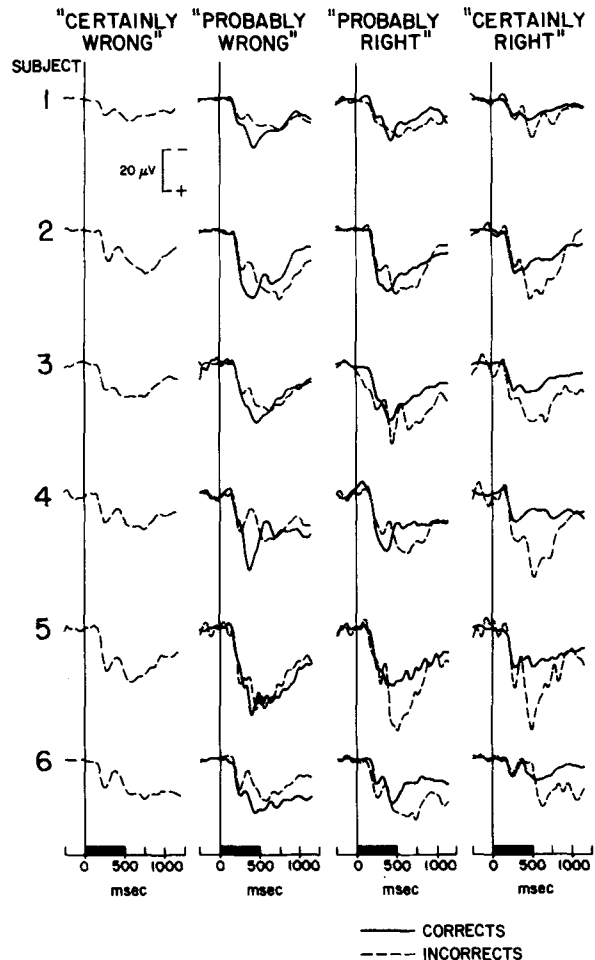


Figure 5. For each subject, the digitally filtered averaged ERPs from Cz that were elicited by the "response" CVC. ERPs from correct and incorrect trials are superimposed for ratings in each of the four confidence ranges. There was an insufficient number of trials in the "certainly wrong"/correct category to consider.

interaction was statistically significant [$F(2,10) = 19.4$, $p < .001$]. When only the "probably wrong" and "probably right" data were analyzed, the Confidence by Outcome interaction remained significant [$F(1,5) = 13.9$, $p < .05$]. A measure of area under the curve (the sum of the digitized voltages between 350 and 950 msec after CVC onset) yielded similar results.

Latency-adjusted P300 amplitude. As stated above, it is possible that the P300 varied considerably in latency from trial to trial. If so, the broad P300s in the average ERPs may not have been representative of the wave-shape on single trials. It is necessary, therefore, to assess the extent to which the apparent amplitude differences observed in the average ERPs might be due to differences in the latency variability of P300 among the single trials that constituted the various averages. To address this question, we latency-adjusted our waveforms, using the adaptive method described by Woody (1967).⁴

Table 3
 $\overline{P300}$ Amplitude Before and After Latency Adjustment for Each Confidence Range by Trial Outcome Category (in Microvolts from Base to Peak)

	Confidence Range			
	CW	PW	PR	CR
Unadjusted Average ERPs				
Corrects		32	26	15
Incorrects	19	26	33	33
Adjusted Average ERPs				
Corrects		36	31	21
Incorrects	21	28	38	40

Note—These amplitudes are grand means over subjects. There were not enough correct trials in the “certainly wrong” confidence range to calculate a valid measure. CW = certainly wrong; PW = probably wrong; PR = probably right; CR = certainly right.

Analyses were done on the single-trial ERPs recorded from Cz, after they were preprocessed with the low-pass digital filter mentioned before. To examine the ERP epoch that contained $\overline{P300}$, the digitized voltages 400-950 msec after CVC onset were analyzed. For comparison with these results, analyses were also performed on an epoch (850-1,500 msec after CVC onset) that presumably contained only background EEG “noise.”

The latency-adjusted average ERPs that resulted from analyses of the $\overline{P300}$ epoch showed slightly sharper $\overline{P300}$ s than did the unadjusted averages. Moreover, the distributions of latencies chosen by these analyses had consistently smaller standard deviations [$F(1,5) = 117.7$, $p < .001$] than did the distributions of latencies chosen by the “noise” epoch analyses. This finding indicates that the Woody (1967) analyses of the epoch containing $\overline{P300}$ indeed detected a latency-varying ERP component, and not simply randomly occurring peaks in the background EEG (see Harris & Woody, 1969).⁵

The mean amplitudes of these latency-adjusted $\overline{P300}$ s, measured on each trial relative to the unadjusted pre-stimulus baseline, are shown for each category in Table 3. An analysis of variance confirmed that the latency-adjusted ERPs manifested the interaction of Confidence Range by Trial Outcome [$F(2,10) = 15.2$, $p < .001$]. Thus the ERP amplitude differences we observed cannot be attributed to differences in the single-trial variability of $\overline{P300}$ latency. Nor can they be attributed to different mixtures of two kinds of ERPs (for example, ERPs with and without a $\overline{P300}$ or ERPs with small vs. large $\overline{P300}$). Distributions of single-trial latency-adjusted amplitudes were examined for each subject and paired associate category. These distributions, summed over subjects, are presented in Figure 6. Bimodal distributions in the categories having large mean $\overline{P300}$ s would have suggested a mixture of two kinds of waveforms. Instead, the $\overline{P300}$ distributions appear to reflect relatively uniform single-trial differences in $\overline{P300}$ among the various categories.

Finally, since both confidence ratings (Table 1) and trial outcome varied with stages of learning, could some variable related to these stages (or to time on task) account for the apparent effect of the interaction of confidence and trial outcome on $\overline{P300}$? Figure 7 shows mean amplitudes of the latency-adjusted averages for combinations of confidence ranges, trial outcomes, and stages of learning. The question here is, when broken down by stages, do trials of each outcome still show $\overline{P300}$ differences related to confidence? Two analyses of variance were performed on the mean latency-adjusted amplitudes, one for correct trials and one for incorrect trials (six subjects with repeated measures on three confidence ranges \times two stages of learning). Both analyses showed statistically significant effects of confidence range [for corrects, $F(2,10) = 25.6$, $p < .001$; for incorrects, $F(2,10) = 8.0$, $p < .01$]. The only other effect that reached the $p < .05$ level of significance was the difference in $\overline{P300}$ amplitude between stages for the incorrects [$F(1,5) = 16.2$, $p < .05$]. Thus, while there was evidence of an effect due to stages of learning, this variable did not account for the interaction of confidence and outcome on $\overline{P300}$.

DISCUSSION

Our data indicate that the amplitude of the $\overline{P300}$ elicited by the “response” CVCs was determined by the

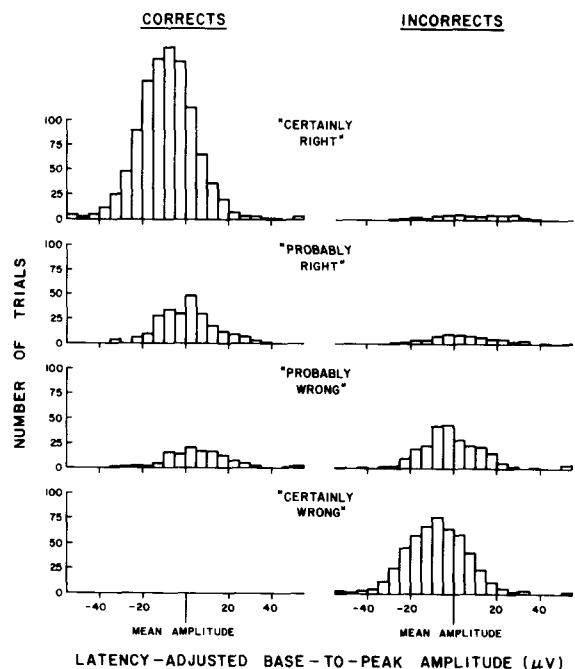


Figure 6. Distributions of the latency-adjusted base-to-peak amplitudes of $\overline{P300}$ on single trials, summed over subjects, for each confidence range and trial outcome category. To adjust for individual differences in amplitude, the mean of each subject's amplitudes over all categories was subtracted from each single-trial amplitude for that subject before the data were combined over subjects.

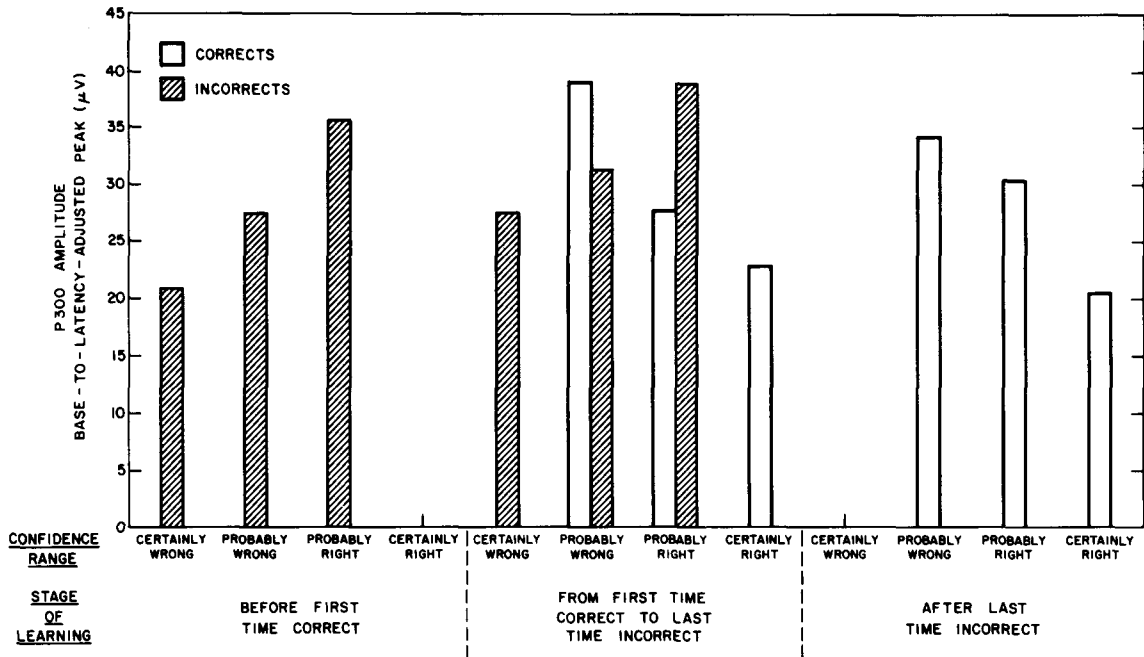


Figure 7. Grand means (over subjects) of latency-adjusted $\overline{P300}$ amplitude. Measures are from Cz ERPs that were elicited by the "response" CVCs on trials from the various combinations of confidence ranges, stages of learning, and correct and incorrect three-letter responses. When broken down by stages, there was an insufficient number of trials in the "certainly wrong"/correct and "certainly right"/incorrect categories.

interaction between a trial's outcome and the subject's expectancy concerning that outcome. Neither confidence by itself nor whether the "response" CVC confirmed or disconfirmed the subject's three-letter response accounted for the variance in $\overline{P300}$. Rather, $\overline{P300}$ amplitude depended on the degree to which the outcome of each trial was unexpected. The lower the subjective probability assigned to the outcome that occurred, the larger was the elicited $\overline{P300}$. These data thus strengthen the claim that $\overline{P300}$ amplitude is dependent on the subjective probability associated with the ERP-eliciting event.

Our notion of subjective probability implies that subjects apply their knowledge about a given situation to form differential expectancies (subjective probabilities) for the various events that might occur. These expectancies, being derived from external information that is filtered by subjects' perceptual biases, stored in a fallible memory, and tainted by an individual's predilections, are "subjective" in that they need not accurately reflect the objective probabilities with which events occur. Information processing triggered by the occurrence of an event is affected by the expectancy associated with that event. An aspect of the processing invoked by unexpected events is reflected in $\overline{P300}$ amplitude. In the paired associate task, it seems reasonable to infer the subjective probabilities that were assigned to "response" CVCs from subjects' confidence ratings. The pattern of these ratings suggests that subjects' confidence accurately reflected their knowledge. To the extent that subjects thought they were correct in the choice of

their three-letter responses, they usually were correct (Figure 2b), and as they learned a list, they indicated more often that they were correct (Table 1).

For the present purposes, it is not necessary to define subjective probabilities rigorously, as one would define mathematical probabilities. We need not, for example, require that the subjective probabilities assigned to all events possible in a given situation sum to one. We need only assume that subjects' expectancies form an ordinal scale. Given that one event is more unexpected than a second event, we predict that the $\overline{P300}$ elicited by the first will be larger than that elicited by the second. Although in some situations (such as the present confidence rating) subjects can articulate their expectations, we do not imply that subjects need necessarily be aware of the probabilities that are internally assigned to events. It seems likely that $\overline{P300}$ is associated, not with the feeling of surprise, but with the processing of surprising events.

In the paired associate task, the events that were assigned differential expectancies were the trial outcomes—information that was derived by subjects from the "response" CVCs. As predicted, when the "response" CVC informed subjects that their three-letter responses were correct, the largest $\overline{P300}$ was elicited if subjects had indicated confidence in the "probably wrong" range. But when the "response" CVC informed subjects that their three-letter responses were incorrect, the largest $\overline{P300}$ was elicited if they had indicated confidence in the "certainly right" or "probably right" range. These trends in $\overline{P300}$ amplitude were evident in

both average and single-trial ERPs. Although both confidence and trial outcome varied as learning occurred, stages of learning could not account for the effects of the interaction of these two variables on $\overline{P300}$ amplitude (see Figure 7). Moreover, since CVC pairs were, overall, presented equally often and were not contingent on either the subjects' three-letter responses or their confidence ratings, the results cannot be due to differences in the frequency with which particular "response" CVCs occurred. Thus, consistent with the results of K. Squires et al. (1973), $\overline{P300}$ was large to the extent that the confidence rating indicated that subjects' expectancies for the obtained trial outcomes were low. Recently, this conclusion was also reached by Campbell, Courchesne, Picton, and K. Squires (1979).

Our results strongly support the suggestion that $\overline{P300}$ reflects the subjective probability for a category to which stimuli are assigned according to the subject's task (see Johnson & Donchin, in press). In one case ("probably wrong"/correct), large $\overline{P300}$ s occurred when the "response" CVC matched the syllable that the subject had presumably activated in memory, having just typed it as the three-letter response; but in other cases ("probably right"/incorrect and "certainly right"/incorrect), large $\overline{P300}$ s occurred when the "response" CVC mismatched the three-letter response. Thus $\overline{P300}$ amplitude was not dependent on whether or not the subject had anticipated the particular "response" CVC that occurred. Rather, the important variable was whether or not the category to which the "response" CVC belonged (denoting correct or incorrect trial outcome) was surprising.

The notion that individuals assign subjective probabilities to events that may occur in the future seems necessary, given the way people deal with uncertainty (see Sheridan & Ferrell, 1974). The less often an uncertain event occurs, the slower subjects respond to it (e.g., Krinchik, 1970), the more reluctant they are to report its occurrence (e.g., Swets, Tanner, & Birdsall, 1961), and the less often they predict that it will occur (e.g., Goodnow, 1955). Much effort has been directed at inferring subjective probabilities from behavioral measures (e.g., Edwards, 1962). In some situations, a normative model provides a reasonable approximation to people's performance in estimating stimulus probabilities and predicting uncertain events (see review by Peterson & Beach, 1967), but systematic biases in subjects' performance reveal that subjective probabilities often do not accurately reflect objective probabilities. Predictions and trial-to-trial estimates of probability are consistently conservative relative to a model of optimal behavior. On the other hand, studies of multistage inference have shown subjects to be too extreme in their probabilistic inferences (see review by Slovic, Fischhoff, & Lichtenstein, 1977). There is also convincing evidence that people sometimes disregard information about probabilities and instead use various heuristics in forming judgments (Tversky & Kahneman, 1974). Further-

more, in a random Bernoulli series, in which successive events are by definition independent, RT responses vary systematically with the sequence of preceding events (see review by Kornblum, 1973).

The present results are consistent with a growing body of evidence that indicates that ERPs also reflect the differential processing of unexpected stimuli. This evidence suggests that the less probable an event is believed to be—whether because it is being presented relatively infrequently (see reviews by Donchin et al., 1978; Ruchkin & Sutton, 1978b), or because it has not occurred recently in a sequence of events (Duncan-Johnson & Donchin, 1977, 1978; Johnson & Donchin, 1979, in press; K. Squires et al., 1976; K. Squires, Petuchowski, Wickens, & Donchin, 1977), or, as shown by the present data, because the event seems unlikely given the subject's current knowledge of a situation—the larger the $\overline{P300}$ will be. Thus when subjective probability varies, $\overline{P300}$ amplitude varies.

The extent to which we can make the converse inference, that events that elicit a larger $\overline{P300}$ are less subjectively probable, depends on the extent to which other variables known to influence $\overline{P300}$ amplitude operate in a given situation. It has been well established, for example, that gradations in the task relevance of an event (Johnson & Donchin, 1978) modulate $\overline{P300}$ amplitude. Indeed, most recent accounts of $\overline{P300}$ have found it necessary to postulate more than one construct in order to explain the systematic variance in $\overline{P300}$ (Donchin, 1979; Donchin et al., 1978; Ruchkin & Sutton, 1978b; K. Squires et al., 1973; Sutton, 1979).

Whether our view of subjective probability is compatible with earlier explanations of $\overline{P300}$ in terms of the resolution of uncertainty and delivery of information (Sutton et al., 1965, 1967) depends on what these two terms are taken to mean. Would more uncertainty be resolved (or more information be delivered) by the "response" CVC when subjects did not think that they knew the appropriate paired associate ("certainly wrong" or "probably wrong") than when they did think that they knew it ("certainly right" or "probably right")? If so, then these constructs do not account for the present results. But if more uncertainty would be resolved or more information delivered on trials having an unexpected outcome, then these conceptualizations seem indistinguishable from that of subjective probability. The importance of the present results is not so much that they argue for the superiority of subjective probability over these other constructs, but that they constrain what must be meant by any construct with which one attempts to account for the observed effects on $\overline{P300}$.

Finally, we emphasize that to relate $\overline{P300}$ amplitude to subjective probability is to assert that $\overline{P300}$ reflects a functional process that is executed differently depending on the subjective probability associated with events. The nature of this process, indeed the functional significance of $\overline{P300}$, remains elusive. At present, some sort

of context-updating operation (see Donchin et al., 1978) seems a likely candidate for the process manifested by $\overline{P300}$.

Knowing the relationship between $\overline{P300}$ and constructs such as subjective probability is useful for integrating past ERP results and for predicting those of future studies. But more important, the relationship suggests that the $\overline{P300}$ provides a family of dependent variables that can be used to augment the analysis of human information processing. We illustrate here the use of $\overline{P300}$ amplitude in assessing subjective probability, but other aspects of $\overline{P300}$, and other endogenous ERP components, might be employed in a similar manner. One advantage of using ERP measures as dependent variables is their multivalued nature. Changes in amplitude, latency, or scalp distribution of the waveform can be used to support inferences about underlying psychological processes. Furthermore, the endogenous ERP components provide direct measures of processes that cannot otherwise be directly observed. Thus, for example, $\overline{P300}$ latency provides a measure of categorization time that is relatively independent of response selection and execution processes (see Donchin & Isreal, in press, for a review). In addition, the ERPs are quite sensitive to the strategies underlying human performance.

It should be emphasized that we are not advocating an abandonment of the traditional tools of cognitive science. We do not believe that cognitive psychologists should trade in their Morse keys for signal averagers. On the contrary, our experience suggests that ERPs yield their best when studied in combination with the overt indices of behavior. We do feel, however, that there are many occasions on which investigators could augment the power of their studies by incorporating ERP measures in their design.

REFERENCES

- AUDLEY, R. J. Some observations on theories of choice reaction time: Tutorial review. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.
- CALLAWAY, E., TUETING, P., & KOSLOW, S. (Eds.). *Brain event-related potentials in man*. New York: Academic Press, 1978.
- CAMPBELL, K. B., COURCHESNE, E., PICTON, T. W., & SQUIRES, K. C. Evoked potential correlates of human information processing. *Biological Psychology*, 1979, **8**, 45-68.
- COURCHESNE, E., HILLYARD, S. A., & COURCHESNE, R. Y. P3 waves to the discrimination of targets in homogeneous and heterogeneous stimulus sequences. *Psychophysiology*, 1977, **14**, 590-597.
- DONCHIN, E. Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked brain potentials and behavior*. New York: Plenum Press, 1979.
- DONCHIN, E., CALLAWAY, E., COOPER, R., DESMEDT, J. E., GOFF, W. R., HILLYARD, S. A., & SUTTON, S. Publication criteria for studies of evoked potentials (EP) in man. In J. E. Desmedt (Ed.), *Attention, voluntary contraction and event-related cerebral potentials*. Basel, Switzerland: Karger, 1977.
- DONCHIN, E., & COHEN, L. Average evoked potentials and intramodality selective attention. *Electroencephalography and Clinical Neurophysiology*, 1967, **22**, 537-546.
- DONCHIN, E., & HEFFLEY, E. Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.
- DONCHIN, E., & ISREAL, J. B. Event-related potentials and psychological theory. In H. H. Kornhuber & L. Deecke (Eds.), *Motivation, motor and sensory processes of the brain: Electrical potentials, behavior, and clinical use*. *Progress in brain research*. Amsterdam: Elsevier-North Holland, in press.
- DONCHIN, E., KUBOVY, M., KUTAS, M., JOHNSON, R., JR., & HERNING, R. I. Graded changes in evoked response (P300) amplitude as a function of cognitive activity. *Perception & Psychophysics*, 1973, **14**, 319-324.
- DONCHIN, E., RITTER, W., & MCCALLUM, C. Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press, 1978.
- DUNCAN-JOHNSON, C. C., & DONCHIN, E. On quantifying surprise: The variation in event-related potentials with subjective probability. *Psychophysiology*, 1977, **14**, 456-467.
- DUNCAN-JOHNSON, C. C., & DONCHIN, E. Series-based vs. trial-based determinants of expectancy and P300 amplitude. *Psychophysiology*, 1978, **15**, 262.
- EDWARDS, W. Subjective probabilities inferred from decisions. *Psychological Review*, 1962, **69**, 109-135.
- FALMAGNE, J. C. Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology*, 1965, **2**, 77-174.
- FRIEDMAN, D., HAKEREM, G., SUTTON, S., & FLEISS, J. L. Effect of stimulus uncertainty on pupillary dilation response and the vertex evoked potential. *Electroencephalography and Clinical Neurophysiology*, 1973, **34**, 475-484.
- FRIEDMAN, D., SIMSON, R., RITTER, W., & RAPIN, I. Cortical evoked potentials elicited by real speech words and human sounds. *Electroencephalography and Clinical Neurophysiology*, 1975, **38**, 13-19.
- FRIEDMAN, D., VAUGHAN, H. G., JR., & ERLNMEYER-KIMLING, L. Stimulus and response related components of the late positive complex in visual discrimination tasks. *Electroencephalography and Clinical Neurophysiology*, 1978, **45**, 319-330.
- GOODIN, D. S., SQUIRES, K. C., HENDERSON, B. H., & STARR, A. An early event-related cortical potential. *Psychophysiology*, 1978, **15**, 360-365.
- GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *American Journal of Psychology*, 1955, **68**, 106-116.
- GRUBIN, M. L., BAUER, J. A., JR., & WALKER, E. C. T. *Alice user's guide*. Natwick, Mass: Alice Associates, 1976.
- HARRIS, E. K., & WOODY, C. D. Use of an adaptive filter to characterize signal-noise relationships. *Computers in Biomedical Research*, 1969, **2**, 242-273.
- HILLYARD, S. A., PICTON, T. W., & REGAN, D. In E. Callaway, P. Tueting, & S. Koslow (Eds.), *Brain event-related potentials in man*. New York: Academic Press, 1978.
- JOHNSON, R., JR., & DONCHIN, E. On how P300 amplitude varies with the utility of the eliciting stimuli. *Electroencephalography and Clinical Neurophysiology*, 1978, **44**, 424-437.
- JOHNSON, R., JR., & DONCHIN, E. Subjective probability and P300 amplitude in an unstable world. *Psychophysiology*, 1979, **16**, 174.
- JOHNSON, R., JR., & DONCHIN, E. P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, in press.
- JOHNSON, R. L., BITZER, D. L., & SLOTTOW, H. G. The device characteristics of the plasma display element. *IEEE Transactions on Electron Devices*, 1971, **18**, 642-649.
- KORNBLUM, S. Sequential effects in choice reaction time: A tutorial review. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press, 1973.

- KRINCHIK, E. P. The probability of a signal as a determinant of RT. In W. G. Koster (Ed.), *Attention and performance II*. New York: Academic Press, 1970.
- KUTAS, M., & DONCHIN, E. Variations in the latency of P300 as a function of variations in semantic categorizations. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.
- KUTAS, M., MCCARTHY, G., & DONCHIN, E. Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 1977, **197**, 792-795.
- LAMING, D. R. J. Subjective probability in choice-reaction experiments. *Journal of Mathematical Psychology*, 1969, **6**, 81-120.
- NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K., & BENT, D. H. *SPSS: Statistical package for the social sciences*. New York: McGraw-Hill, 1975.
- NOBLE, C. Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, **8**, 487-521.
- PETERS, J. F., BILLINGER, T. W., & KNOTT, J. R. Event related potentials of brain (CNV and P300) in a paired associate learning paradigm. *Psychophysiology*, 1977, **14**, 579-585.
- PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29-46.
- RITTER, W., SIMSON, R., & VAUGHAN, H. G., JR. Association cortex potentials and reaction time in auditory discrimination. *Electroencephalography and Clinical Neurophysiology*, 1972, **33**, 547-555.
- ROTH, W. T., FORD, J. M., & KOPELL, B. S. Long-latency evoked potentials and reaction time. *Psychophysiology*, 1978, **15**, 17-23.
- RUCHKIN, D. S., & GLASER, E. M. Some simple digital filters for examination of CNV and P300 waveforms on a single trial basis. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978.
- RUCHKIN, D. S., & SUTTON, S. Emitted P300 potentials and temporal uncertainties. *Electroencephalography and Clinical Neurophysiology*, 1978, **45**, 268-277. (a)
- RUCHKIN, D. S., & SUTTON, S. Equivocation and P300 amplitude. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potential research* (EPA-600/9-77-043). Washington, D.C.: U.S. Government Printing Office, 1978. (b)
- SHERIDAN, T. B., & FERRELL, W. R. *Man-machine systems: Information, control, and decision models of human performance*. Cambridge, Mass: M.I.T. Press, 1974.
- SLOVIC, P., FISCHHOFF, B., & LICHTENSTEIN, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, **28**, 1-39.
- SMITH, S., & SHERWOOD, B. A. Educational uses of the Plato computer system. *Science*, 1976, **192**, 344-352.
- SQUIRES, K. C., DONCHIN, E., HERNING, R. I., & MCCARTHY, G. On the influence of task relevance and stimulus probability on event-related potential components. *Electroencephalography and Clinical Neurophysiology*, 1977, **42**, 1-14.
- SQUIRES, K. C., HILLYARD, S. A., & LINDSAY, P. H. Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. *Perception & Psychophysics*, 1973, **13**, 25-31.
- SQUIRES, K. C., PETUCHOWSKI, S., WICKENS, C., & DONCHIN, E. The effects of stimulus sequence on event related potentials: A comparison of visual and auditory sequences. *Perception & Psychophysics*, 1977, **22**, 31-40.
- SQUIRES, K. C., WICKENS, C., SQUIRES, N. K., & DONCHIN, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, 1976, **193**, 1142-1146.
- SQUIRES, N. K., DONCHIN, E., SQUIRES, K. C., & GROSSBERG, S. Bisensory stimulation: Inferring decision-related processes from the P300 component. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 299-315.
- STUSS, D. T., & PICTON, T. W. Neurophysiological correlates of human concept formation. *Behavioral Biology*, 1978, **23**, 135-162.
- SUTTON, S. P300—Thirteen years later. In H. Begleiter (Ed.), *Evoked brain potentials and behavior*. New York: Plenum Press, 1979.
- SUTTON, S., BRAREN, M., ZUBIN, J., & JOHN, E. R. Evoked-potential correlates of stimulus uncertainty. *Science*, 1965, **150**, 1187-1188.
- SUTTON, S., TUETING, P., ZUBIN, J., & JOHN, E. R. Information delivery and the sensory evoked potential. *Science*, 1967, **155**, 1436-1439.
- SWETS, J. A., TANNER, W. P., JR., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301-340.
- TUETING, P., SUTTON, S., & ZUBIN, J. Quantitative evoked potential correlates of the probability of events. *Psychophysiology*, 1970, **7**, 385-394.
- TVERSKY, A., & KAHNEMAN, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, **185**, 1124-1131.
- WOODY, C. D. Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and Biological Engineering*, 1967, **5**, 539-553.

NOTES

1. In keeping with conventional paired associate terminology, the words "stimulus" and "response" refer to the CVCs displayed by the computer. The CVC typed by the subject is called here the "three-letter response."

2. Many of the 101 points on the confidence scale were not used by a given subject, whereas other points were used quite often. Some points, by themselves, were used more than 4% of the time. Points 0 and 100 were the ratings used most often by all subjects. Note that the structure of the task predisposed a large proportion of ratings to the upper end of the scale. Since CVC pairs were presented at random until all were learned, subjects received many presentations of pairs that they already knew before they received a sufficient number of presentations of the pairs that they did not yet know.

3. To avoid confusion in referring to ERP components that can vary in latency, we have adopted the notation suggested by Donchin, Callaway, Cooper, Desmedt, Goff, Hillyard, and Sutton (1977). The positive-going wave that occurs here at a modal latency of 280 msec is denoted P280. The late positivity, thought to be the same entity that in some previous experiments occurred at 300 msec, but which occurs here at a much longer latency, is denoted the P300.

4. The Woody procedure, which has been used previously in ERP work (Kutas et al., 1977; Ruchkin & Sutton, 1978a), involves calculating the cross-correlation function between each single-trial waveform and a template of the ERP signal that varies in latency. The lag at which the maximal cross-correlation (or, as used here, cross-covariance) occurs is assumed to be the latency of the signal on that trial. The single-trial ERPs can then be shifted relative to each other to time-lock on the signal, and a latency-adjusted average can be computed. We used two different approaches to derive templates of the latency-varying signal. First, templates were derived by an iterative procedure (Woody, 1967) whereby the latency-adjusted average of one iteration served as the template for the next iteration, this sequence proceeding until the template stabilized. The unadjusted average ERP served as the template for the first iteration. These analyses, since they were done for each paired associate category and subject separately, were sensitive to any differences that might have existed in the waveshape of the latency-varying ERP component among the various categories. But since there were fewer trials in some categories than others, the reliability of the various derived templates might have differed

systematically across categories. Therefore, to derive a single template that could be applied to each of the seven paired associate categories, we took advantage of evidence (presented in the text) that the late positivities seen in both the counting and the paired associate tasks were composed of the same component: $\overline{P300}$. Thus, as a second approach, we latency-adjusted each subject's counted CVC ERPs and used this average as the template for a one-pass cross-covariance analysis of single ERPs from each of that subject's paired associate categories. Since both approaches yielded the same pattern of statistically significant results, we report only the results of the iterative analyses. The analyses of the "noise" epochs also followed the iterative approach, but the latency-adjusted ERP elicited by the counted CVC was for each subject used as the template for the first iteration.

5. This result does not imply that the detection of $\overline{P300}$ was equally reliable in all categories. In the categories in which $\overline{P300}$ amplitude was relatively small, the Woody procedure may have chosen a spurious EEG peak on a larger proportion of

trials than in the categories in which $\overline{P300}$ was relatively large. Consistent with this possibility, the standard deviations of latencies tended to be largest in the categories in which $\overline{P300}$ amplitude was smallest. For using the Woody procedure to confirm differences among categories in $\overline{P300}$ amplitude after adjusting for latency variability, such trends are not problematic. But these trends do make it difficult to draw conclusions about systematic differences in the latency of $\overline{P300}$, since mean latencies could be biased by the proportion of spurious trials chosen in the various categories. Thus while the waveforms in Figure 5 show apparent differences in $\overline{P300}$ latency as well as in amplitude, whether these latency differences reflect an overall effect of trial outcome or an interaction between outcome and confidence remains obscure.

(Received for publication November 19, 1979;
revision accepted April 9, 1980.)