AUDITORY EVOKED-RESPONSE DIFFERENTIATION WITH DISCRIMINATION LEARNING IN HUMANS

DAVID JENNESS

Biometrics Research, New York State Department of Mental Hygiene, New York, New York, and Columbia University

The possible functional significance of the registration of the salience of stimulus events in the vertex potential was investigated with a discrimination-learning paradigm involving stimuli that were equally informational but of different value. Wave forms were obtained over the entire course of acquisition. Level of discriminative accuracy and extent of evoked-response differentiation were closely related, and differentiation had to predominantly reflect improvement in the perceptual-cognitive "labeling" of the stimuli. When the identical stimuli served a feedback rather than a cueing role, evoked responses were markedly different, and there was no apparent relation to accuracy or learning.

The sensory evoked response recorded from scalp in human subjects reflects parameters of the physical stimulus used; perhaps more important, late components of the averaged response reflect relatively complex psychological variables. Early investigators reported research involving variables such as vigilance, arousal, and voluntary attention, with the predominant finding being that increases along these dimensions are associated with greater amplitude of various components of the response, particularly the large biphasic (negative-positive) wave known as the vertex potential (Davis, 1964). Generally, such studies depend on the manipulation of the subject's "state" by means of instructions or an intrinsic task feature like difficulty or complexity. Sometimes, however, the argument is circular, in that evoked-potential differences are offered as the only evidence that this control over the subject's state was achieved. (Critical reviews of experiments in this area are given by Donchin and Cohen, 1967; Jenness, 1971; Karlin, 1970; Näätänen, 1967; Tueting, 1969.)

Experiments of this kind are unlikely to tell us much about the information-carrying properties, if any, of evoked responses—that is, about the possibility that the evoked-response wave form, in part or as a whole, can reflect complex psychological processing with regard to the signal value of a stimulus. This is so especially when the analysis of the electrocortical response is not related in some way to specifiable behavioral outcomes. In certain respects, Uttal's (1965, 1967) question of whether the cortical evoked potential is a sign or a code is relevant here. Recently, several researches have established that the appearance of "late" components in the averaged evoked potential is a function of stimulus meaningfulness or salience, and that these components represent the effect of the subjects' past experience, their attitude toward the stimulus events, or their assessment of stimulus importance. In the human realm, Sutton, Braren, Zubin, and John (1965) have called these components "endogenous"; in animals, a similar conception is expressed by the phrase "read-out from memory" (John, 1967).

Beginning in 1965, Sutton and colleagues (Sutton et al., 1965; Sutton, Tueting, Zubin, & John, 1967; Tueting, 1969; Tueting, Sutton, & Zubin, 1970) have demonstrated that a large late positive component of the vertex evoked potential, \( P_3 \) (about 300 msec. from stimulus onset), is present when the stimulus delivers information and absent or

---

1 This research was partially supported by Grants MH-14580 and MH-07997 from the National Institute of Mental Health. The author is deeply grateful to Samuel Sutton for his support and advice during the conduct of this research and during its preparation for publication.

2 Requests for reprints should be sent to David Jenness, who is now at the Social Science Research Council, 230 Park Avenue, New York, New York 10017.
greatly reduced in amplitude when the stimulus is redundant. Donchin and Cohen (1967) interpreted the occurrence of the late positive component in the visual evoked potential in terms of "the degree to which a stimulus delivers task relevant information to the subject [p. 545]." Ritter, Vaughan, and Costa (1968) reported that P3 appeared to the occurrence of stimulus events that represented a change in stimulation that could not be predicted by the subject; these authors interpreted their findings as reflecting "perceptual orienting responses"—which could amount to the reaction to an influx of relevant "information." (A similar interpretation has been provided by Klinke, Fruehstorfer, and Finkenzeller, 1968.)

However—and here we point to an aspect of theoretical development, not a lack of ingenuity or methodological precision—such experiments still treat the subject essentially as a passive receiving instrument for stimulation. Salience of one kind or another is attributed to, or discovered in, stimulus events by an observer. What has not yet been demonstrated is that such stimulus-delivered salience is functional. Ideally, one would wish to establish that reliably different evoked responses are directly associated with distinct behavioral outcomes. Pribram, Spinelli, and Kamback (1967), in monkeys, and John and colleagues (John & Killam, 1960; John, Shimokochi, & Bartlett, 1969), in cats, have approximated such demonstrations in discrimination-learning experiments with animals, though the electrocortical responses involved, and certain other aspects, are not directly comparable to those in human studies to date.

Gardiner (1969) found differences in the auditory evoked response when subjects were engaged in a pitch discrimination task as opposed to a loudness discrimination task, but did not analyze his electrophysiological results in terms of psychophysical performance. Cross-sectional studies of discrimination performance (as opposed to discrimination-learning studies) have been reported by Davis (1964), Garcia-Austt, Bogacz, and Vanzulli (1964), Haider, Spong, and Lindsley (1964), and Ritter and Vaughan (1969), but the first three studies are more naturally interpreted as vigilance or selective-attention studies. Ritter and Vaughan (1969) imposed visual or auditory discrimination tasks on subjects in a situation where, on certain trials, a flash or tone burst was diminished in intensity. The data show large late positive components to detected signals (i.e., stimulus change), but not to undetected signals or nonsignals. However, their subjects apparently gave "miss" reports (to undetected signals) but not "false alarm" reports (to nonsignals as if they were signals). Thus, in a situation where no trial-by-trial rewards or penalties were imposed, the subjects must have been operating with a cautious psychophysical bias; and a precise basis for relating the various categories of electrocortical response to the behavioral (psychophysical) categories in the experiment is lacking. Ritter and Vaughan (1969) reported also that when for some subjects the discrimination task was made extremely difficult the late positive component occurred in each outcome category, persisting even when the discriminative performance became quite good. In a discrimination experiment by Hillyard, Squires, Bauer, and Lindsay (1971), subjects performing near threshold provided all four categories of response, but showed the late positive component only to correctly reported signals.

The present study attempted to monitor in the human evoked potential a slow incremental learning process, by obtaining wave forms over the entire course of acquisition in a situation where the consequences of the responses were themselves distinctly different. From the literature, it could be predicted that the wave forms would include prominent late components, reflecting the subjects' assessment of stimulus identity and meaning. The task was designed to be very difficult, since it was hoped that a gradual differentiation of evoked responses during the course of the experiment could be shown to be closely related to the precision with which the task was performed.
Method

Subjects and Apparatus

Four university students, aged 18–23 yr., served as subjects. They were seated in a dimly lit, copper-screened, grounded chamber facing a curved white screen on which brief light flashes from a fluorescent source were thrown and in the center of which a small fixation-point light bulb was mounted. For responding, subjects were provided with two red-switch buttons that closed with minimal force with the movement of the index finger. Clicks were presented through the left side of a pair of Sharp e HL-10 earphones set in circumaural cushions. The clicks were transients passed through Model 330M Krohn-Hite filters so as to achieve a low click (1,000–2,000 Hz.) or a high click (1,250–2,000 Hz.), the two clicks being balanced to yield equal amplitude at amplifier output (earphone input), which corresponded to about 60 db. re sensation level. Effective click duration was about 3 msec. Further specifications are given in Jenness (1971).

Three Beckman silver/silver-chloride electrodes were used, at midline frontal, vertex, and occipital locations on the scalp. For all derivations, an electrode at the right earlobe (contralateral to the stimulated ear) provided a common monopolar reference. Miniature Beckman electrodes were used at earlobe and at a location 1 cm. below the left eyelid, which provided an eye-to-ear channel for recording eye-movement artifacts (Hillyard, 1968; Low, 1966; Vaughan, 1969). Scalp-to-ear resistances were kept below 5,000 ohms at all times.

Electroencephalographic (EEG) and electrooculographic (EOG) signals were led out of the chamber into four matched Model P511 ac Grass preamplifiers (three EEG, at 10,000× gain; one EOG, at 5,000× gain); these were set to achieve half-amplitude response at .15 and 100 Hz. The EEG channels premiated frontal, vertex, and occipital derivations from the midline; the EOG channel recorded vertical eye movements. A fifth data channel involved the use of a long-time-constant ac-coupled Dynamics Instrumentation Co. Model 890 dc amplifier, used in its manual mode so that its automatic resetting feature was eliminated. The effective half-amplitude band pass lay between .0165 and 100 Hz. in this channel, which was used to record both sensory evoked-potential activity and any slower steady-potential changes from the vertex location. Since comparisons of data recorded on the two vertex channels revealed that the averaged evoked-potential activity was represented equally well in both channels, but that the Grass amplifier failed to reflect very slow shifts, the vertex data to be reported here are all obtained from the lower frequency (.0165–Hz.) channel.

Amplified EEG and EOG activity was recorded with the use of a Sangamo Model 3562 FM tape recorder. Recording was done at 1% in/sec, playback at 60 ips, with better than 1% speed-transfer accuracy. The appropriate trials, segregated by channel and experimental category, were later fed into a summing computer, the Fabriette FT 1052 computer of average transients. Each memory address of the computer integrated graded input amplitude from tape over 6.41 msec. of real time. The summed voltages were normalized, and the averaged wave form was written out on graph paper using a Moseley Model 7030 AM X–Y plotter.

Procedure

Figure 1 provides a diagram of an experimental trial. The experimenter initiated each trial according to a semirandom sequence that determined which click stimulus would occur. The exact moment of trial onset, however, was outside his control, in order to prevent inadvertent pacing of the trials. The first trial event, a 3-msec. flash coincident with fixation-light onset, told the subject to fixate and get ready. After 1.6 sec., the “task” click—one or the other discriminative stimulus—occurred. After another 1.6 sec., a second flash told the subject to respond—to push one of his two buttons and record his judgment. Two seconds thereafter the same click that had served as the task click on that trial occurred again if the subject had been correct. If the subject had given the wrong response, the final click did not occur. The offset of the fixation light terminated the trial.

Fairly long time segments were recorded prior to, between, and following the click stimuli that evoked the more important evoked responses, so that EEG samples without stimulus-referred activity were available for inspection. Furthermore, using a respond cue to delay the subject’s overt response made it less likely that motor or premotor activity should contaminate evoked-potential segments.

Before the actual experiment was begun, each subject was pretrained on the general procedure in order to train out eye movements and to familiarize him with the task. (Two clicks different from the clicks subsequently used, but of comparable discriminative difficulty, were used in the pretraining.) It was explained to the subject that while the click stimuli were objectively of different frequency compositions, the perceived basis for judgment could be a cue other than pitch. “You are to use whatever difference you hear... we’re really asking you to put the right label, the right name, to the clicks.” The subject was also instructed as follows:

In addition to your hourly pay, we are going to pay you a bonus according to your performance. Every time you say “high” you win 5¢ if you were right—if the click in fact was high. You lose 5¢ if you say “high” and are wrong. We call the high response the “hot” response, because when you make it you win or lose money. The response “low” is a neutral response; when you say “low” you neither win nor lose. With the bonus, by the end of the
Fig. 1. Diagram of the experimental trial. (The \( t_0 \) refers to the onset of the trial; it does not coincide with the beginning of the averaging epoch.)

...experiment you can make as much as $10 or $15 a day extra.

Alternatively, another subject would be instructed that his response "low" would win or lose money and that "high" was the neutral response. Finally, the subject was told:

If you think about it, you'll see that the bonus doesn't change the task. You still have to label the clicks. If you responded "high" to every click, you would win 5¢ half the time, but half the time you would lose 5¢. If you were very cautious and said "low" all the time, you would never lose money—but you'd never win any, either. Understand?

The task clicks were presented in blocks of 50 trials, each block taking a little less than 10 min. Ample between-block and midsession rest periods were provided. The trial protocol was such that in 300 trials each click occurred 150 times. Between 450 and 600 trials were obtained each session, depending mainly on the subject's morale.

When eye movements possibly contaminating EEG activity were seen on the oscilloscope monitor, the trial was erased without comment by the experimenter. Well-trained subjects produced few such trials. Erased trials were always replaced at the end of the programmed block, in the order in which they had been erased.

Sessions took place about twice a week. Between 9 and 11 experimental sessions were conducted with the subjects. Each subject's participation was terminated when a high level of accuracy had been reached, and/or when accuracy seemed to have reached a plateau, was declining, or seemed to be becoming more variable.

In the case of two subjects the final session was run with the discriminative stimuli made more difficult. Here, the low-click band pass remained 1,000–2,000 Hz, while the high-click band pass was changed to 1,200–2,000 Hz. The two clicks were balanced for energy in the usual way.

**Principles of Data Analysis**

**Composition of averages.** The contingent relation between stimulus events and judgments of stimulus identity can be represented as shown in Figure 2.

The experimental situation is such that discrimination learning is reflected in the tendency for trials to accumulate, to a greater and greater degree, in the hits and correct rejections cells. It is feasible to analyze the data across sessions by sorting on the stimulus. When analyzed longitudinally over the course of the experiment, changes in column-averaged evoked-potential wave forms should reflect some changing aspect of stimulus judgment or evaluation. If the two averaged evoked responses are different but no changes in these averages are seen while discriminative performance changes, the evoked response must reflect stimulus...
input primarily in its physical aspect, irrespective of stimulus labeling on the part of the subject.

It is feasible also to analyze the data by sorting on the response. That an analysis by rows should also reflect evoked-potential changes associated with discrimination learning is perhaps less obvious. However, all readouts, no matter how sorted, represent stimulus-evoked potentials that precede the behavioral response and that occur while the subject evaluates the nature of the stimulus occurrence with respect to the response contingencies in force. The nature of the experiment is such that the salience of the (perceived) stimulus is response defined, while the occurrence of gradually improving discriminative accuracy is in itself evidence that discriminative responses are coming under the control of the cue properties of the stimuli. If row-derived averaged evoked responses are different, but fail to change with changes in discrimination performance, this could indicate that their essential characteristics were determined merely by the subject's assessment of the nature of the impending response (e.g., on the basis of behavioral "risk"), irrespective of stimulus characteristics or perceptual evaluation.

It is also technically possible to obtain four readouts, representing four cells, as primary data, and to make a number of comparisons among them. This approach to analysis, however, raises some problems. As discriminative performance improves, the number of hits and correct rejections must increase, while the number of false alarms and misses must decrease, and this is likely to continue beyond the point where computer averaging is reliable. However, properties of these wave forms can be algebraically estimated from knowledge of the corresponding quantities from the hits or correct rejections readout and a column or row readout.

We generally report evoked-potential difference quantities between columns, rows, or cells of the analytic matrix. Results based on differences reduce the data and the number of comparisons, and facilitate relating the electrophysiological to the behavioral results (i.e., single values for psychological accuracy). The analysis of column and row differences can ensure that "pure" stimulus or response effects, as discussed hypothetically above, do not predominate in the data. (If they do, differences along a marginal should yield something approaching a constant, regardless of accuracy.) Finally, difference measures seemed appropriate for interpretive purposes in an experiment where the main issue was whether discrimination learning would be accompanied by differentiation of stimulus-evoked responses recorded cortically.

Other procedures. In order to take into account variables that determine subjects' behavior over and above those effects due to stimulus properties, it was decided to express accuracy of judgment with respect to either stimulus in terms of the number of hits or correct rejections given the distribution of responses in the row containing that cell. The following formula yielded a single accuracy index for a given session:

\[ \% \text{ correct after chance} = \left( \frac{\% \text{ hits} - 50\%}{\% \text{ correct rejections} - 50\%} \right) \]

Figure 2 shows an example of a full-epoch, 6.56-sec. averaged evoked-potential vertex readout. The three large rectangular waves are calibration pulses. Arrows indicate the time of presentation of stimuli, except that in the case of the feedback (second) click the arrow indicates the moment of occurrence or nonoccurrence of that click. For technical reasons, in full-epoch readouts the apparent amplitude of the feedback-click response is artificially small, since the averaging n is that of the column or row involved. This is not true of that response in half-epoch readouts, from which all sensory evoked-potential measures were made. The measures were obtained by drawing base lines covering 131 msec. prior to stimulus onsets, thus permitting measurements of response components in terms of the distance of their peaks and troughs from this arbitrary zero line. In this report, however, peak-to-peak (or peak-to-trough) measures are used. Full-epoch readouts were used to evaluate slow-potential shifts bridging stimulus events, the possible presence of artifacts or time-locked alpha activity in "empty" parts of the epoch, etc.

The identification of components in the averaged evoked response was based on considerations of latency, direction of change, and shape as reported in the human literature (for components up to \( P_3 \)), and on systematic latency data collected from each subject in each session, from which medians, ranges, and typical latencies were determined (see Jennett, 1971). In some cases—for example, the late component \( P_4 \), whose relative diminution in certain conditions was of interest—the assigned amplitude value was the most extreme point in the wave form within a 50-msec. range around the typical latency.

**Results**

Across experimental sessions, subjects improved the accuracy with which they identified the two task clicks. These behavioral results, in terms of percentage correct after chance, are shown in Figure 4. Subjects showed considerable variability of performance from day to day at higher accuracy
levels. Nonetheless, the overall ascending nature of the accuracy curves, especially over the first four or five sessions in all subjects, indicates that discrimination learning did occur. This conclusion is supported, in Subjects LR and JV, by the drop in accuracy when the discriminanda were made more difficult.

Inspection, across days, of subjects’ response distributions (e.g., the ratio of hot to neutral responses) did not suggest that any one aspect of response bias or strategy—for example, cautious vs. risky—was associated with the main accuracy trends or with electrophysiological measures.

**Task-Click Evoked Responses**

In Figure 5, examples are displayed of vertex-derived averaged evoked potentials obtained from the task clicks. These are
Fig. 5. Superimposed task-click evoked potentials. (Left column: accuracy less than 40% after chance; right column: accuracy greater than 60% after chance. Solid trace: hot stimulus wave form; broken trace: neutral stimulus wave form. Vertical broken lines mark moment of stimulus onset. Time is in milliseconds. Negative-going components appear as upward pen deflections.)
wave forms from stimulus-sorted trials; they are column averages. In each panel of the figure, the solid line is the hot wave form, while the broken line is the neutral. They were superimposed by making their prestimulus base lines coincide. For each subject, a pair of wave forms is shown on the left from an early session, where accuracy was less than 40% (after chance), and another pair on the right from a later session, where accuracy was greater than 60%.

Figure 5 suggests certain essentially qualitative generalizations. At fairly low levels of accuracy, slight amplitude differences in the wave forms can be seen in most subjects in some of the components. These differences occur in either negative-going or positive-going components. At higher levels of accuracy, wave-form differences are larger, the growth in magnitude of these differences being associated primarily with components later than P2. From early to later performance levels, wave-form differences seem to change in magnitude, not in kind. It appears that individual components do not emerge or disappear or take on radically new shapes. Certain overall trends—for example, generally greater positivity in later sessions—are true of both wave forms, hot and neutral, but to different degrees. Subjects differ in the relative size of their own wave-form components, in the magnitude and location of differences evinced in the wave form, and in the degree to which these increase from early to late performance. Generally, however, the P3/N3/P4 segment of the wave form accounts for most of the differential longitudinal change.

**Individual component analysis.** Peak-to-peak amplitudes for the N1/P2, P2/N2, N2/P3, P3/N3, and N3/P4 components were plotted across sessions, in separate hot and neutral stimulus-averaged plots for each subject. Inspection of these double plots indicated similar longitudinal trends across subjects in terms of changes in the size and prominence of components. Standard evoked-potential components could be recognized and measured across each subject’s sessions; adaptation or habituation across sessions was not evident; and the later components tended to increase in size over sessions. The data were then plotted as wave-form differences (always hot minus neutral) across sessions, component by component and subject by subject. The results indicated, for all subjects, that for Components N2/P3 and P3/N3 the constant direction of the difference quantity tended to be that of hot > neutral, and that the difference quantity tended to increase in the later sessions. Components P2/N2 and N3/P4 tended, though to a lesser degree, to show an opposite trend—that is, the direction of difference being neutral > hot, with this difference also increasing in magnitude across sessions. Component N1/P2 showed no consistent trend.3

A more exact analysis of the specific relationship between evoked-potential differences and level of discriminative performance would seem to involve two issues. One is, As accuracy improves, do evoked-potential differences, component by component, increase? The answer here was provided by computing the Spearman rank-order correlations, rs (Siegel, 1956), between percentage correct and the magnitude of difference, for each component in each subject. The other issue is, As accuracy changes from day to day, does the evoked-potential difference change in a systematic way? This specific relationship, which deals only with pairs of adjacent points, will be called “tracking.” The number of times the evoked-potential differences increased as accuracy increased or decreased as accuracy decreased (positive tracking) was computed, and this was expressed as a ratio in respect of the number of comparisons. For the correlations, the X data points (see Figure 4) were excluded, while for the computation of tracking, they were included.

A comprehensive matrix of correlation coefficients and tracking ratios revealed only scattered values substantially different from the expected values under a null hypothesis of rs = 0.00 for the correlations and P (proportion) = .5 for the tracking measure. No single subject showed consistently high values on either measure, nor did any single component. However, except for

---

3Details of the analysis of individual components are given in Jenness (1971).
N1/P2, the obtained values tended to range in one direction or another from the expected values.

Composite measures. If evoked-potential differences as measured in individual components in individual subjects are small, and do not reliably reflect a differentiation of hot and neutral wave forms, a combination of these effects might yield a more substantial and stable difference. The component-by-component results, however, showed that the direction of difference in N2/P3 and N3/P4 was opposite to that in N2/P3 and P3/N4. Hence, the combined difference was maximized rationally by performing a linear combination of hot-related minus neutral-related component differences, beginning with P2, with the P2/N2 and N3/P4 quantities entering into the combination negatively.

The results of this combination of stimulus-sorted task-click differences are plotted in Figure 4, where the curve for each subject may be compared directly to his accuracy curve. Since sums of differences each of whose directions with respect to zero are truly random ought themselves to be randomly positive and negative, the fact that out of 38 evoked-potential data points 36 are positive in sign is highly improbable owing to chance. And under the null hypothesis, it remains true with respect to the composite that the expected measures of rank-order correlation and of tracking are $r_{s} = 0.00$ and $P = .5$, respectively. The obtained measures are shown in the top line of Table 1. The correlation between level of discrimination accuracy and magnitude of the composite difference is significant in all four subjects. The tracking measure is strongly positive in three of these four subjects; in the fourth subject, the tracking ratio is precisely at its chance value.

The positive tracking relationships demonstrated in the data of Subjects JF, DW, and LR are due equally to an increase in the composite difference when accuracy increased and a decrease when accuracy decreased (see Figure 4). It is not, for example, the case that tracking was due merely to an increasing overall trend in both measures, so that it is but a by-product of the rank-order correlation.

The same analyses were performed on task-click differences based on row-derived (response-sorted) averaged evoked-response wave forms. The same components were identified and measured as in the stimulus-sorted data, and the same composite measure was obtained for all subjects. As shown in the second line of Table 1, the results were entirely compatible, and with respect to tracking virtually identical, with those based on column differences. This

<table>
<thead>
<tr>
<th>Evoked-response quantity</th>
<th>Correlation coefficients</th>
<th>Tracking ratios*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot minus neutral (stimulus)</td>
<td>JF</td>
<td>DW</td>
</tr>
<tr>
<td>Hot minus neutral (response)</td>
<td>.85**</td>
<td>.67*</td>
</tr>
<tr>
<td>Hits minus correct rejections</td>
<td>.84**</td>
<td>.85**</td>
</tr>
<tr>
<td>Hits</td>
<td>.94**</td>
<td>.65*</td>
</tr>
<tr>
<td>Correct rejections</td>
<td>.29</td>
<td>.92**</td>
</tr>
<tr>
<td>False alarms minus misses</td>
<td>.58</td>
<td>.01</td>
</tr>
<tr>
<td>False alarms</td>
<td>-.04</td>
<td>-.05</td>
</tr>
<tr>
<td>Misses</td>
<td>-.47</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td>-.54</td>
<td>-.03</td>
</tr>
</tbody>
</table>

* The denominator is the number of tracking comparisons. The numerator represents the number of times the task-click difference increased across 2 adjacent days when percentage correct increased, or decreased when accuracy decreased, plus one-half the number of ties—cases where the task-click difference remained the same across 2 days.

** $p < .05$.

* $p < .01$. 

TABLE 1

Relation of Task-Click Evoked-Potential Composite Measures to Discriminative Accuracy in Various Trial Categories.
supports the prediction that stimulus-evoked potential differentiation would develop during the course of learning, and that the effect would be apparent when the basis for trial sorting was either the actual identity of the stimulus or the nature of the discriminative response. Taken together, the two sets of results indicate that the obtained stimulus-sorted results cannot be attributed, for example, to an increasing number of hot response trials falling, across sessions, into a hot stimulus cell, thus yielding characteristics that are fundamentally response determined—an interpretation that would be indicated if the row-difference (“pure” response) result across sessions were a constant, or approximately so. Conversely, the change with learning reflected in the response-sorted results cannot be attributed to more and more neutral stimulus trials, for example, falling into the neutral response row, thus contributing “pure” stimulus characteristics regardless of labeling on the part of the subject; were such the case, the hot minus neutral stimulus quantity would be a constant across sessions, reflecting differential brain response to the stimuli as physical or sensory (i.e., prelabeled) entities.

The task-click evoked-response data for the successful outcome trials, hits and correct rejections, were analyzed separately, following the data-reduction stages described earlier. The results, based on hit minus correct rejection difference curves in relation to accuracy, are also given in Table 1. Tracking ratios are in general unaltered, and the rank-order correlations are comparable to those obtained from row or column differences. The one exception is JV, who in this purified averaged evoked-response difference measure now shows neither significant tracking nor correlation with accuracy. Aside from JV’s data, however, the relationships obtained for stimulus– or response-sorted trials hold for successful-outcome trials. In particular, this analysis shows that the tracking aspect of column or row differences is not due merely to fluctuations across sessions in the homogeneity of the pooled trials entering into the average (i.e., with these categories becoming relatively purer as accuracy increases).

The same particularization of the data along the false alarm/miss diagonal—error trials—yields quite a different result. The algebraic subtraction across sessions of the misses from the false alarms quantity in the four subjects gives results that show neither orderly tracking nor significant rank-order correlations. These curves oscillate irregularly and quite widely around the null-difference line. Whatever wave-form differences may exist between these two categories of error trials, they are apparently not stable differences, nor does the extent of such a difference vary with accuracy. (See Table 1.)

The fact is, as Figure 5 and the description of steps in data reduction may suggest, that the various task-click wave forms did not in an obvious way fall into sets of qualitatively different categories corresponding to the analytic categories employed. Differences, while consistent, were small in size, complex, and distributed across a number of components. It can be seen, for example, in the correlation data of Table 1 that the significant hits minus correct rejections differences for JF, DW, and LR rest on quite different contributing relationships. For example, JF’s correct rejections–accuracy correlation is apparently negative, unlike that between hits and accuracy, while for LR both are strongly positive. What seems to be the common characteristic of all these subjects is that the hits–accuracy correlation is more positive than the corresponding one involving correct rejections. The tracking ratios show a similar overall pattern in correct-outcome relationships; again, JF, DW, and LR vary as to whether the correct rejections ratio is about at chance or shows an inverse relation to accuracy. The essentially random relationships based on false alarms minus misses rest, in the cases of JF, DW, and LR, on constituent relations to accuracy that are either weak, lie around a chance level, or are close together in value.

These are three subjects, then, where the hits and correct rejections are well differentiated, while the error trials are not. These data may support the conclusion that for evoked-response differentiation and discriminative accuracy to progress together.
certain characteristics of the electrocortical responses associated with hits and correct rejections must diverge, while those associated with errors must not—or, at least, must not behave like either of the correct-outcome responses.

Other Results

Figure 6 shows clearly that the feedback click, which was physically identical to the task click, evoked markedly different potentials. (The trial category involved for the comparison is that of hits; correct rejections data give a similar result.) The amplitude of the initial negative–positive component was from about 30% (DW) to about 100% (JV) larger than in the task click. This was due primarily to the depth of the positive process; in fact, the N1 component was much smaller than its task-click counterpart in most subjects. The large trough-to-peak upward component ending at about 400 msec was, except for LR, two to four times as large as any similar upward-going process in the task-click potentials. The wave forms of Subject JV showed only these two large components reliably. The wave forms of the other subjects sometimes showed two inner components, an upward and a downward one, but they were small and irregular in appearance. Of the subjects with inner components, only JF consistently showed a positive trough around 300 msec. that was as large or larger than that around 200 msec.; in the other subjects, the earlier positive process was larger and more clear-cut.

Evoked-potential differences were measured component by component, and then were summed across the four (or two) components and, alternatively, across only the two outer components in all subjects. There was no consistent relation between accuracy and the size or the direction of the differences between the hot- and neutral-related components (as determined by the response to the task click) or their composites, either in terms of overall trends or session-to-session changes. In DW, LR, and JV the averaged wave form obtained from the hot feedback stimulus (hits) was generally larger in amplitude than that evoked by the neutral feedback stimulus (correct rejections), but in JF the opposite was true. A prediction on our part that the feedback-click wave-form differences would be maximal at low levels of discrimination accuracy and would diminish thereafter was not supported. Nor did the data support our expectation that, with increasing accuracy of judgment of the task click, the size of the feedback-click response would diminish. Nor was there any consistent association between day-to-day changes in the feedback-click difference curves and the composite difference curves for the task click.

As for slow-potential shifts between or following the stimulus events of the trial (see Figure 3), subjects again showed activity that was fairly reliable within, but different across, subjects. All subjects showed a small negative shift between the occurrence of the task click and the flash commanding the response, after which a large positive shift developed. This positive process was rapid and deep, reached an inflection or “corner” at about 500–600 msec. after the flash, and terminated in a plateau that stayed positive (i.e., did not drift back toward the earlier base-line level) during the approximately 1,000 msec. remaining before the delivery of the feedback click.

The negative shift following the task click was uniformly small—generally about 8 \( \mu \text{V} \). (measured from the base line prior to the task click to the base line prior to the respond flash), and never exceeding 12 \( \mu \text{V} \). The size of the shift, averaged across hot and neutral stimuli, tended to decline across days. However, the precise shapes of these generally declining curves did not reflect any consistent association—based on either rank-order or tracking measures—with accuracy, feedback-click evoked-response size, or size of the task-click evoked-response composite difference. Various analyses were made, but there was no evidence for a differentiation in form or magnitude developing between the negative shift following hot stimuli and that following neutral stimuli as discriminative performance improved.

No aspect of the positive shift that occurred with or following the response was clearly related to any other aspect of the
Fig. 6. Evoked-potential wave forms, from the same subjects, to physically identical stimuli serving two different functions in the experimental procedure. (Wave forms shown begin at moment of stimulus onset, but were superimposed according to average microvolt level during the 131 msec. prior to onset. Left column: accuracy less than 40% after chance; right column: accuracy greater than 60% after chance.)
data in the study. In the evoked-potential epoch as a whole, no interrelationships were found among distinct aspects of the electrophysiological data, variously elicited in the several segments of the trial. An analysis was undertaken, on a fairly exhaustive level, of possible interrelationships among accuracy, task-click, feedback-click, negative-shift, and positive-shift measures for hot and neutral averages separately, in case difference measures or mean-size measures might have obscured true associations. No such associations were found.

In addition, certain “control” analyses were undertaken in order to test the validity of the results. The “empty” segments of the evoked-potential wave forms, hot and neutral, were always compared visually to make sure that averaged EEG activity was essentially of the same order and kind. Averaged EOG activity was compared in each set of data obtained; all records were free of timed eye-movement artifact, except at the very beginning of the epoch (where fixation was achieved) and at the end of the trial, when reflex eye movements often occurred, but sufficiently distant in time from feedback-click onset that the click-evoked response itself was not affected. The amplitude of the early negative (occasionally positive) shift following the first flash and preceding the occurrence of the task click, the hot-minus-neutral differences associated with the averaged early-shift amplitudes, and the mean size and hot-minus-neutral differences in $N_1/P_2$ of the first flash-evoked potential were plotted to ensure that no systematic preclick differences conceivably due to differential arousal or expectation could explain subsequent task-click evoked-response differences. There was no evidence for any such systematic factors.

The mean $N_1/P_2$ amplitudes of the first warning-flash evoked potentials, the task-click evoked potentials, and the feedback-click evoked potentials (here, $N_1/P_{200}$) for each subject were plotted as parallel curves across sessions in order to assess possible epoch-wide changes related to adaptation or habituation effects. None of these measures was reliably related to accuracy or to the task-click composite difference, and the sets of three curves showed no apparent intra-subject associations. This suggests that arousal or adaptation factors were generally uniform throughout the trial epoch and throughout the course of the experiment in each subject.

In frontal recordings, the task-click and feedback-click responses were grossly like those obtained from vertex, but the across-session relation of the former to discrimination accuracy did not hold. Auditory evoked responses as recorded from the occipital location were too small and ill-defined for reliable measurements to be made.

**Discussion**

**Task-Click Results**

The demonstration of evoked-potential differentiation with discrimination learning depended on the readout of the response-determined or response-relevant label, while the psychophysical results—orderly changes in response distributions across sessions—make it clear that the nature of the response became more and more a function of stimulus properties. Thus, as in instrumental learning experiments generally, stimulus and response come to be defined in terms of each other, as distinct behavioral sequences are elaborated. The differentiated wave forms obtained in this study must reflect the fact that two distinct behavioral units were formed.

The rank-order correlation and the tracking measure together provide persuasive evidence for the proposition that the level of performance and the nature and extent of a difference in the averaged evoked responses to hot and neutral stimuli are directly related. If the correlation had been high, but with very poor tracking being shown, a reasonable alternative interpretation would have been that the evoked-potential curve reflected a motivational gradient over the experiment as a whole. On the other hand, the two curves might have shown perfect point-to-point tracking relationships, but the evoked-potential curve could have merely crossed and recrossed a constant horizontal level, so that the rank-order correlations would be very low or nil. This
kind of result could have been attributed to an either/or kind of emotional or motivational reaction on the part of the subject.

JV's task-click results present something of a problem in interpretation: his stimulus- and response-sorted evoked-response differences are significantly related to discrimination accuracy, while his tracking and other measures are not. However, JV's accuracy curve (Figure 4) is relatively flat and is of a shape that suggests that discrimination learning in this subject was unstable and subject to retrogressions. The chance-level tracking measure may reflect this subject's inability to perform the discriminative task reliably. Therefore, in this subject as well as the others, behavioral and electrophysiological measures may in fact be closely related.

The task-click results suggest that something more may be involved in the evoked response than merely the occurrence of appropriately labeled "read-out from memory," in John's (1967) phrase. The accuracy of discrimination in a given session and the magnitude of the evoked-response difference in that session seem to be directly associated, to vary together. This may imply that the evoked responses are not just correlates of perceptual activity but represent, reflect, or provide for a processing or coding mechanism that may in part determine which response is emitted. This involves the conception that different evoked responses carry functionally different information.

In this interpretation, the present results may seem at variance with those obtained by Ritter and Vaughan (1969) in humans and by John and colleagues (John & Killam, 1960; John et al., 1969) in cats, in that in those data evoked activity in error-outcome behavioral categories could be more readily interpreted. For example, in the Ritter and Vaughan paper it was noted that misses "looked like" correct rejections. The present experiment investigated the development of certain brain-behavior relationships longitudinally, and is not ideally designed to examine "typical" wave shapes corresponding to the four analytic cells cross-sectionally, in terms of similarity or dissimilarity. The failure to find stable differences in misses vs. false alarms here may be due to the fact that the wave forms in these two cells are different in some way that does not lend itself to the present method of measurement; or that the small n in these cells makes measurement or estimation inherently inaccurate; or that the averaged activity in each of the two cells is essentially random in nature, so that the subtraction of one from the other yields nothing of interest. The last of these possibilities would suggest that these cells may contain trials on which the subject fails to attain a clear percept; hence, the evoked responses are either extremely heterogeneous or in some other way unclear.

A rather similar interpretation, centering on the possible role of subjective certainty, has been made by Hillyard et al. (1971), in presenting data where only hits showed the late positive component.

Further, once learning has been demonstrated and evoked-response measures are claimed to reflect or depend on the course or stage of learning, there is a theoretical objection that can be made to the notion that in all cases distinct evoked-response wave forms should be sufficient to predict a specific behavioral response. Briefly, if the electrocortical activity in an error-outcome trial should be exactly like that in a correct-outcome trial, when both correspond to the same behavioral report, how does learning ever take place? It is not tenable that learning should progress when the identical covert response is sometimes reinforced (albeit remotely) and sometimes not reinforced. Other models of learning, involving the notion of the brain's looking at its own activity or its trace, encounter similar difficulties. It is perhaps important that in the experiments of Ritter and Vaughan (1969) and John et al. (1969), either asymptotic performance without articulated reinforcement contingencies (in the former study) or generalization performance, which is a kind of extinction situation, was involved. Brain-behavior relationships in long-established or overlearned tasks may differ from those apparent during acquisition.

Various segments of the evoked-response wave form could, of course, reflect or pro-
vide different “codes.” The fact that differences around P₃—or around the large “late positive component” (Ritter & Vaughan, 1969)—tend most clearly to differentiate the wave forms in this study is consonant with most human evoked-potential research with meaningful or significant stimuli (see Ritter & Vaughan, 1969; Sutton et al., 1965; Sutton et al., 1967; Tueting, 1969). Systematic differences in the N₃/P₄ component are, it is believed, obtained here for the first time. The actual occurrence of a P₄ component may be due to the fact that the experimental task involved pitch discrimination, since Gardiner (1969) reported that, at least in some subjects, a positive-going process of about this latency was present in a tone–evoked wave form when the subjects discriminated on the basis of pitch, but not on the basis of intensity. Further research needs to be undertaken in order to assess the properties of P₄.

The present results are generally in keeping with the interpretation of the late positive-going evoked-potential process (here, N₂/P₃/N₃) in terms of stimulus salience, but there are differences in components later than those reported by Sutton and colleagues or by Donchin and Cohen (1967), and the interpretation here must have a different character from that of Tueting et al. (1970), who interpreted the amplitude of P₃ as directly related to the probability of the occurrence of a stimulus. Note that the loose term “salience” seems necessary for the interpretation of the present results, especially. Strict conceptualizations involving stimulus information cannot apply here: From the point of view of information theory, for example, both stimuli were equally informational—but one was more salient than the other.

Although the manner in which late positive components were obtained by Ritter and Vaughan (1969) was very different from the way in which the present components were obtained, the present data seem to be in conflict with their interpretation of the component as reflecting an unpredictable but meaningful change in stimulus input. An orienting response or related notion would hardly suggest that the late positive component or components should increase in magnitude, absolutely and differentially, over the course of the experiment and as a function of accuracy of performance. It is not reasonable to suppose that the present subjects found the stimuli more unpredictable in later sessions.

In summary, the present task-click results constitute perhaps the most compelling evidence to date from humans for the proposition that evoked responses differentiated gradually during the course of instrumental learning demonstrate the formation of information-carrying brain mechanisms that are not trivially associated with the occurrence of differentiated responses to which tangible consequences are attached. It is hoped that future human evoked-potential research will test this assertion, since at present it has primarily speculative status.

Feedback-Click Results

The size and shape of the feedback-click responses here are not so notable in themselves as the way in which, in the same subjects in the same experiment, the effect of stimulus role has been so strikingly demonstrated. Given the fact that the feedback-click wave form as an entity seems different from the task-click wave form, what becomes of the distinction between an earlier, exogenous, evoked portion of the response and a later, endogenous, elicited portion? Assuming, as one is led to, that the feedback click is highly “salient” and informational, where are the prominent late components (e.g., P₃) that are quantitatively related to the degree of stimulus significance? Is the larger simplified shape of the response characteristic of reinforcing stimuli (here, secondarily reinforcing) or—a somewhat different conception—of stimuli that deliver task-relevant information very quickly and compellingly?

The present data seem to make less likely Ritter and Vaughan’s view (1969) that, except when a difficult discrimination is not involved—as it was not, presumably, in the case of the feedback click here—a late positive component occurs only if the stimulus elicits an orienting response. Rather, the
present results seem to indicate that there may be several or many vertex auditory evoked potentials, depending on the functional role of the stimulus as well as on certain other stimulus properties and experimental operations. This speculation encourages investigation along two lines, either of which could be important from the psychological or physiological point of view.

Certain classes of stimuli and of experimental operations may turn out to produce distinct types of evoked response, whose characteristics are to some extent sui generis. Research to date may have sampled only a few of these. On the other hand, it may turn out that investigators to date have sampled evoked responses too widely and discontinuously, and need now empirically to fill in hitherto excluded middle ground. Further research may result in a logical and quantitative ordering of experimental operations and a corresponding calculus for the evaluation of evoked responses that will render the idea of distinct evoked-response types meaningless.

REFERENCES


NÄÄTTÄNEN, R. Selective attention and evoked potentials. Annals of the Finnish Academy of Science (Suomakaisen Tiedaskatemian Toimitukset), 1967, 151, 1–226. (Separate)


(Received August 4, 1971)